

Testing General Relativity with 21 cm intensity mapping

Alex Hall,^{1,*} Camille Bonvin,^{1,2} and Anthony Challinor^{1,2}

¹*Institute of Astronomy and Kavli Institute for Cosmology, Madingley Road, Cambridge, CB3 0HA, U.K.*

²*DAMTP, Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, U.K.*

We investigate the prospects for constraining alternative theories of gravity with a typical near-term low-budget 21 cm intensity mapping experiment. We derive the 21 cm brightness temperature perturbation consistently in linear theory including all line-of-sight and relativistic effects. We uncover new terms that are a small correction on large scales, analogous to those recently found in the context of galaxy surveys. We then perform a Fisher matrix analysis of the B_0 parametrization of $f(R)$ gravity, where B_0 is proportional to the square of Compton wavelength of the scalaron. We find that our 21 cm survey, in combination with CMB information from Planck, will be able to place a 95% upper limit of 7×10^{-5} on B_0 in flat models with a Λ CDM expansion history, improving on current cosmological constraints by several orders of magnitude. We argue that this constraint is limited by our ability to model the mildly non-linear regime of structure formation in modified gravity. We also perform a model-independent principal component analysis on the free functions introduced into the field equations by modified gravity, μ and Σ . We find that 20–30 modes of the free functions will be ‘well-constrained’ by our combination of observables, the lower and upper limits dependent on the criteria used to define the ‘goodness’ of the constraint. Our analysis reveals that our observables are sensitive primarily to temporal variations in Σ and scale variations in μ . We argue that the inclusion of 21 cm intensity maps will significantly improve constraints on any cosmological deviations from General Relativity in large-scale structure in a very cost-effective manner.

I. INTRODUCTION

A major challenge in cosmology is understanding the late-time accelerated expansion of the Universe [1, 2]. If we assume that gravity is described by General Relativity (GR), such an expansion requires the existence of a new exotic form of energy with negative pressure, called dark energy. An alternative solution is to give up GR at large (cosmological) scales and introduce modifications to the laws of gravity (for a review, see [3]).

Whilst observations that measure the expansion history can rule out specific dark energy and modified-gravity models, there is usually sufficient freedom in the form of the modification to reproduce any desired expansion history, thus rendering these measurements unable to distinguish between a (possibly time-dependent) dark energy component and modified gravity [4]. To constrain these models, we must therefore go beyond the level of the background, and study the perturbed Universe. It is here that an unclustered dark energy component may be distinguished from modified gravity¹.

Current and future optical surveys focus mainly on two promising observables: galaxy number counts and cosmic shear. These observables are both sensitive to the distribution of matter in the Universe and therefore provide powerful probes of the laws of gravity. In this paper we concentrate on a complementary observable measured in radio surveys: the redshifted 21 cm emission line of neutral atomic hydrogen (HI). We are interested in the post-reionization emission that traces the distribution of hydrogen at low redshifts. The advantage is that after reionization, the physics of the 21 cm power spectrum is simple and can be accurately modelled in linear theory [7, 8]. This is in contrast to observations of the epoch of reionization, where the 21 cm power spectrum is very sensitive to the details of the reionization process [9].

The post-reionization signal is thought to be dominated by emission from dense self-shielded damped Ly α systems (DLAs) which form in high-density regions, with some of the signal originating from optically-thin Ly α absorbers in low-density regions (for reviews, see [10–12]). On large (linear) scales, it is expected that these sources trace out the underlying matter density field, up to a scale-independent bias factor if the underlying density field is Gaussian (for non-Gaussian corrections to this argument, see for example [13]). The 21 cm post-reionization signal therefore provides a new window to observe the distribution of matter. Recently, measurements of the 21 cm cosmological signal at $z = 0.8$ have been made by cross-correlating with optical galaxies [14, 15].

*Electronic address: ach74@ast.cam.ac.uk

¹ Note however that some modified-gravity models may still remain indistinguishable from a clustered dark energy model at the perturbed level [5, 6].

The main advantage of the 21 cm method over standard optical surveys is that it does not require the resolution of individual galaxies: one detects the diffuse line radiation from a large number of sources. This technique, called *21 cm intensity mapping* [16] thus allows one to observe large volumes of the sky efficiently with low-resolution interferometers or single-dish experiments which can be constructed relatively cheaply [17–19]. It also permits observation at higher redshifts than optical surveys, which are limited by the necessity to identify galaxies individually, a task that becomes increasingly difficult as galaxies appear fainter. A further advantage of observing the 21 cm line is that source redshifts are readily obtainable from the observation frequency. This allows one to perform a tomographic analysis with very thin redshift bins, and consequently to study the evolution and scale dependence of the underlying matter density field. Moreover, precise redshift information enhances the significance of the redshift-space distortion term, which is suppressed in optical surveys with only photometric redshifts by uncertainties in the redshifts. It follows that 21 cm surveys are very well suited to cosmological parameter inference [20], measurement of baryon acoustic oscillations (BAO) [21, 22], constraining the properties of dark energy [23], and constraining models of modified gravity [24, 25].

Experimentally, observing at low redshifts (i.e. after reionization) has the advantage that the foreground contamination from Galactic synchrotron emission is significantly reduced compared to epoch-of-reionization experiments, since its amplitude scales roughly as $\nu^{-2.7}$ (for a thorough investigation of foregrounds in the context of a single dish experiment, see [17]). Although foreground removal will still present a significant challenge in recovering the cosmological signal, this should be possible since the foreground components are smooth in frequency space, as opposed to the signal which has significant structure in the frequency (i.e. radial) direction due to the discrete nature of the underlying sources and their small-scale clustering.

However, there is a price to pay for observing the Universe at high frequencies. Since most of the HI is destroyed during the epoch of reionization, the signal is significantly smaller at post-reionization redshifts. Although the foregrounds are also smaller as mentioned above, the system temperature of the detector is still significant due to the receiver noise floor. A typical antenna temperature for a near-term experiment is 50 K, while the sky temperature at low redshifts is typically 5–10 K [11]. The system temperature at the receiver is thus dominated by the design of the antenna. A further disadvantage of working with the low redshift signal is that the typical co-moving length scale below which linear perturbation theory is expected to break down is larger than at high redshift, due to hierarchical structure formation. Since we do not attempt to model the non-linear 21 cm signal in this work, the number of independent modes of the density field accessible to us is much smaller than at high redshift. However, even with these considerations, the 21cm intensity mapping method provides a potentially powerful tool for learning about late-time structure formation, at a fraction of the cost of spectroscopic galaxy surveys or high-redshift epoch-of-reionization 21 cm surveys such as those planned with the Square Kilometre Array (SKA).

In this paper, we provide a consistent derivation of the perturbation to the 21 cm brightness temperature in linear theory, including all relativistic effects which could be relevant on large angular scales, analogous to the new terms found in the context of galaxy surveys in [26–28]. Our forecasts of the potential of 21 cm mapping for constraining modified-gravity models complements the work of Masui et al. [24], which considered only information from the BAO signal and the reconstructed weak lensing signal in two specific models. In contrast, we include the full three-dimensional clustering information in angular and redshift space provided by the intensity maps and consider more general modifications to gravity. We try to keep our experimental setup as general as possible, but choose a redshift range of $0.7 \leq z \leq 2.5$, appropriate to that of the Canadian Hydrogen Intensity Mapping Experiment (CHIME) [18].

The plan of this paper is as follows. In Sec. II we derive the mean 21 cm brightness temperature and its perturbation consistently in linear theory, including all line-of-sight and relativistic effects. In Sec. III we discuss models of modified gravity, and the response of the 21 cm signal to deviations from GR. We also set up the parametrizations of modifications to gravity there which are used for our forecasts. In Sec. IV we discuss 21 cm intensity mapping and the experimental setup assumed, before describing our statistical methods and presenting our results in Sec. V. We conclude in Sec. VI.

The cosmological parameters varied in our statistical analyses are $(\Omega_b h^2, \Omega_c h^2, \tau, h, A_s, n_s, \Omega_\nu h^2)$, with fiducial values of (0.0223, 0.112, 0.085, 0.73, 2.04×10^{-9} , 0.966, 0.0006) and pivot scale 0.05 Mpc^{-1} for the primordial power spectrum. The background expansion history is always that of Λ CDM models, and we assume three neutrino species with degenerate masses. Our metric convention is $+, -, -, -$ and, unless stated otherwise, we take $c = 1$.

II. THE 21 CM BRIGHTNESS TEMPERATURE

In this section we derive the perturbation to the 21 cm brightness temperature in linear theory, including all relativistic effects. Much of the material in this section is drawn from [29] and Appendix A of [27]. Except where stated explicitly, the results in this section are valid for all metric theories of gravity.

Let the rest-frame (proper) number density of neutral hydrogen atoms at redshift z along some line of sight be n_{HI} , with a fraction n_1/n_{HI} being in the excited triplet states and n_0/n_{HI} in the singlet state of the 21 cm hyperfine

transition. Neglecting the finite line width of the emission, which should be fine on the large scales we consider, in the rest-frame of the gas the net change in the number of photons per volume with energy E and $E + dE$ propagating within a solid angle $d\Omega$ in proper time dt due to 21 cm interactions is

$$dn_{\text{emit}} = \frac{1}{4\pi} [(n_1 - 3n_0)\mathcal{N}_\gamma + n_1] A_{10}\delta(E - E_{21})dEdtd\Omega. \quad (1)$$

Here, \mathcal{N}_γ is the photon occupation number, $A_{10} \approx 2.869 \times 10^{-15} \text{ s}^{-1}$ is the spontaneous emission coefficient [30], and $E_{21} = 5.88 \mu\text{eV}$ is the rest-frame energy of a 21 cm photon. In Eq. (1), we have assumed that the atomic triplet states are populated isotropically (see [29] for further discussion of this point).

The level populations define the spin temperature T_s by $n_1/n_0 = 3e^{-T_{21}/T_s}$, where $T_{21} \equiv E_{21}/k_B = 0.068 \text{ K}$ and k_B is Boltzmann's constant. We assume that the radiation field at the relevant frequencies consists of the CMB blackbody with temperature $T_{\text{CMB}} \gg T_{21}$ and the additional photons from the 21 cm interaction. At low redshift, $T_s \gg T_{\text{CMB}}$ [10] since the spin temperature is coupled to the gas temperature by $\text{Ly}\alpha$ photons, and stimulated emission and absorption can be neglected². In this limit, Eq. (1) becomes

$$dn_{\text{emit}} \approx \frac{3}{16\pi} n_{\text{HI}} A_{10} \delta(E - E_{21}) dEdtd\Omega. \quad (2)$$

Note that this result is independent of the spin temperature.

Since re-absorption is negligible, and neglecting Thomson scattering of the anisotropies in the line radiation (the Thomson optical depth to $z = 2$ is 0.008), we can calculate the brightness temperature by simply summing up the emitted photons. These are received by an observer with 4-velocity u^a along a line-of-sight direction $\hat{\mathbf{n}}$. In terms of the photon distribution function $f(E, \hat{\mathbf{n}})$, the number of photons collected with energies between E and $E + dE$ in an area dA subtending a solid angle at the observer of $d\Omega$ in proper time dt is

$$dn_{\text{rec}} = f(E, \hat{\mathbf{n}}) E^2 dEd\Omega dAdt. \quad (3)$$

We relate dn_{rec} and dn_{emit} by considering the propagation of the congruence of null geodesics that focus at the observer. Let $d\tilde{A}$ be the invariant area of the wavefront at affine parameter λ corresponding to some source position. In an interval of affine parameter $d\lambda$, the wavefront sweeps out a volume $d\tilde{A} u_s^a k_a d\lambda$, where u_s^a is the source 4-velocity and k^a is the wavevector, $k^a = dx^a/d\lambda$. If the collecting area of the detector, dA , subtends a solid angle $d\tilde{\Omega}$ at the source in its rest-frame, then photons in an (observer-frame) energy range dE around energy E will be collected in time dt in solid angle $d\Omega$ with number

$$\begin{aligned} dn_{\text{rec}} &= \frac{3}{16\pi} \int d\lambda \left[n_{\text{HI}} A_{10} \delta[E(1+z) - E_{21}] (1+z) dE \frac{dt}{1+z} d\tilde{\Omega} d\tilde{A} k_a u_s^a \right] \\ &= \frac{3}{16\pi} n_{\text{HI}} A_{10} (1+z) \left| \frac{d\lambda}{dz} \right| dEdtd\tilde{\Omega} d\tilde{A}, \end{aligned} \quad (4)$$

where the integral in the first equation is along the line of sight, and in the second the quantities are evaluated at redshift z along the line of sight, where $1+z = E_{21}/E$. (Additionally, we used $k_a u_s^a = E_{21}$.) Using the reciprocity relation $d\tilde{A} d\tilde{\Omega} = dA d\Omega / (1+z)^2$, comparison with Eq. (3) gives

$$f(E, \hat{\mathbf{n}}) = \frac{3}{16\pi} \frac{n_{\text{HI}} A_{10} (1+z)}{E_{21}^2} \left| \frac{d\lambda}{dz} \right|. \quad (5)$$

Finally, the 21 cm brightness temperature T_b is related to the photon distribution function by $k_B T_b = h^3 E f / 2$, where h is Planck's constant. It follows that

$$T_b(z, \hat{\mathbf{n}}) = \frac{3}{32\pi} \frac{h^3 n_{\text{HI}} A_{10}}{k_B E_{21}} \left| \frac{d\lambda}{dz} \right|. \quad (6)$$

If we initially ignore perturbations, $|dz/d\lambda| = (1+z)H(z)E_{21}$ (where $H(z)$ is the Hubble parameter) and the background brightness temperature is (reinstating factors of c)

$$\begin{aligned} \bar{T}_b(z) &= \frac{3(hc)^3 \bar{n}_{\text{HI}} A_{10}}{32\pi k_B E_{21}^2 (1+z) H(z)} \\ &= 0.188 \text{ K } h \Omega_{\text{HI}}(z) \frac{(1+z)^2}{E(z)}, \end{aligned} \quad (7)$$

² This follows for the CMB photons since $T_s \gg T_{\text{CMB}}$. For the additional 21 cm photons, the optical depth for re-absorption can be shown to be $O(10^{-2})T_{21}/T_s \ll 1$ around $z = 1$, and is therefore negligible.

where $E(z) = H(z)/H_0$, the Hubble constant $H_0 = 100h \text{ kms}^{-1}\text{Mpc}^{-1}$, and $\Omega_{\text{HI}}(z)$ is the *comoving* mass density in HI in units of the *current* critical density. Where required in this work, we take $\Omega_{\text{HI}} = 4 \times 10^{-4}$ to be constant, consistent with the local value found in the HIPASS survey [31]. The value of Ω_{HI} is not expected to vary significantly over the redshift range we consider [32].

In the presence of perturbations, in Eq. (6) we must evaluate the perturbed n_{HI} at the perturbed position appropriate to the redshift z and line of sight $\hat{\mathbf{n}}$, and include the perturbation in $|dz/d\lambda|$. We will work in the conformal Newtonian gauge, where the metric takes the form

$$ds^2 = a^2(\eta) [(1 + 2\Psi)d\eta^2 - (1 - 2\Phi)\delta_{ij}dx^i dx^j]. \quad (8)$$

Here, η is conformal time, $a(\eta)$ is the scale factor and the spacetime-dependent gravitational potentials are Ψ and Φ .

An observer at rest in this coordinate system has 4-velocity $u^\mu = a^{-1}(1 - \Psi)\delta_0^\mu$, which is parallel to ∂_η . We can equip this observer with an orthonormal tetrad of 4-vectors, $(e_\alpha)^a$ such that $(e_0)^a = u^a$ and $e_i = a^{-1}(1 + \Phi)\partial_i$. Decomposing the wavevector onto the tetrad gives

$$\frac{d\mathbf{x}}{d\eta} = (1 + \Phi + \Psi)\mathbf{e}, \quad \frac{d\eta}{d\lambda} = a^{-2}\epsilon(1 - \Psi), \quad (9)$$

where the comoving energy ϵ is defined by $k_a u^a \equiv \epsilon/a$, and the unit 3-vector \mathbf{e} consists of the spatial tetrad components of the propagation direction. In addition we have the geodesic equations for the photons

$$\frac{d\mathbf{e}}{d\eta} = -\nabla_\perp(\Phi + \Psi), \quad \frac{d\epsilon}{d\eta} = -\epsilon \frac{d\Psi}{d\eta} + \epsilon(\dot{\Phi} + \dot{\Psi}), \quad (10)$$

where $\nabla_\perp \equiv \nabla - \mathbf{e}\mathbf{e} \cdot \nabla$. The derivatives here are along the line of sight and overdots denote partial derivatives with respect to η .

The source velocity (i.e. the bulk velocity of the HI) may be written as $u_s^a = u^a + v^a$ such that $u_s^\mu = a^{-1}[1 - \Psi, v^i]$ where v^i are the spatial tetrad components of v^a . Similarly, the 4-velocity of an observer at position A moving with 3-velocity v_{oA}^i in the conformal-Newtonian frame can be written $u_{oA}^\mu = a_A^{-1}[1 - \Psi_A, v_{oA}^i]$. The redshift of the source measured by this observer at A is then

$$1 + z = \frac{a_A}{a} \frac{\epsilon}{\epsilon_A} (1 + \hat{\mathbf{n}} \cdot [\mathbf{v} - \mathbf{v}_{oA}]), \quad (11)$$

where $\hat{\mathbf{n}} = -\mathbf{e}$ is the line-of-sight direction of the photons as seen by the observer. (Aberration and evolution of \mathbf{e} can be neglected to the order we are working.) Integrating Eq. (10) gives the ratio of Newtonian-gauge energies

$$\frac{\epsilon}{\epsilon_A} = 1 + \Psi_A - \Psi + \int_{\eta_A}^{\eta} (\dot{\Phi} + \dot{\Psi}) d\eta', \quad (12)$$

and so the redshift is

$$1 + z = \frac{a_A}{a(\eta)} \left(1 + \Psi_A - \Psi + \int_{\eta_A}^{\eta} (\dot{\Phi} + \dot{\Psi}) d\eta' + \hat{\mathbf{n}} \cdot [\mathbf{v} - \mathbf{v}_{oA}] \right). \quad (13)$$

Writing the perturbed conformal time at redshift z along the line of sight $\hat{\mathbf{n}}$ as $\eta(\hat{\mathbf{n}}, z) = \bar{\eta}_z + \delta\eta$, where $\bar{\eta}_z$ is the unperturbed value, it follows from Eq. (13) that

$$\mathcal{H}(\bar{\eta}_z)\delta\eta = \Psi_A - \Psi + \int_{\eta_A}^{\bar{\eta}_z} (\dot{\Phi} + \dot{\Psi}) d\eta' + \hat{\mathbf{n}} \cdot [\mathbf{v} - \mathbf{v}_{oA}], \quad (14)$$

where all quantities on the right-hand-side are evaluated on the zero-order lightcone. Here, $\mathcal{H} \equiv \dot{a}/a$ is the conformal Hubble parameter.

Writing the perturbed n_{HI} as $\bar{n}_{\text{HI}}(1 + \delta_n)$ and evaluating at $\eta(\hat{\mathbf{n}}, z)$, the density term in Eq. (6) becomes $\bar{n}_{\text{HI}}(\bar{\eta}_z)[1 + \delta_n + (\dot{\bar{n}}_{\text{HI}}/\bar{n}_{\text{HI}})\delta\eta]$. For $|d\lambda/dz|$, we differentiate Eq. (13) and use Eq. (9); evaluating at the perturbed conformal time gives

$$\left| \frac{d\lambda}{dz} \right| (z, \hat{\mathbf{n}}) = \frac{a(\bar{\eta}_z)}{\mathcal{H}(\bar{\eta}_z)E_{21}(1+z)} \left[1 - \left(\frac{\dot{\mathcal{H}}}{\mathcal{H}} - \mathcal{H} \right) \delta\eta - \frac{1}{\mathcal{H}} \frac{d\Psi}{d\eta} + \frac{1}{\mathcal{H}} (\dot{\Phi} + \dot{\Psi}) + \frac{1}{\mathcal{H}} \hat{\mathbf{n}} \cdot \frac{d\mathbf{v}}{d\eta} + \Psi + \hat{\mathbf{n}} \cdot \mathbf{v} \right]. \quad (15)$$

Combining these results, and using $\mathcal{H} = aH$, gives the background brightness temperature in Eq. (7). The fractional perturbation to the brightness temperature is³

$$\Delta_{T_b}(z, \hat{\mathbf{n}}) = \delta_n + \frac{\dot{n}_{\text{HI}}}{n_{\text{HI}}} \delta\eta - \left(\frac{\dot{\mathcal{H}}}{\mathcal{H}} - \mathcal{H} \right) \delta\eta - \frac{1}{\mathcal{H}} \frac{d\Psi}{d\eta} + \frac{1}{\mathcal{H}} (\dot{\Phi} + \dot{\Psi}) + \frac{1}{\mathcal{H}} \hat{\mathbf{n}} \cdot \frac{d\mathbf{v}}{d\eta} + \hat{\mathbf{n}} \cdot \mathbf{v} + \Psi. \quad (16)$$

If the comoving number density of HI is conserved at low redshift (i.e. the ionized fraction of hydrogen is constant), $\dot{n}_{\text{HI}}/\bar{n}_{\text{HI}} = -3\mathcal{H}$. Using the Euler equation $\dot{\mathbf{v}} + \mathcal{H}\mathbf{v} + \nabla\Psi = 0$,⁴ some further cancellations occur leaving

$$\Delta_{T_b}(z, \hat{\mathbf{n}}) = \delta_n - \frac{1}{\mathcal{H}} \hat{\mathbf{n}} \cdot (\hat{\mathbf{n}} \cdot \nabla\mathbf{v}) + \left(\frac{d \ln(a^3 \bar{n}_{\text{HI}})}{d\eta} - \frac{\dot{\mathcal{H}}}{\mathcal{H}} - 2\mathcal{H} \right) \delta\eta + \frac{1}{\mathcal{H}} \dot{\Phi} + \Psi. \quad (17)$$

Equation (17), new to this work, is the expression we have been seeking for the 21 cm brightness temperature perturbation in the limit of zero frequency bandwidth. It is straightforward to show that it agrees with the more complete expression [Eq. (18)] in [29] in the limit $T_s \gg T_{\text{CMB}}$ and for low optical depth for re-absorption and Thomson scattering.

Each of the terms on the right of Eq. (17) has a simple physical explanation. The first two terms are the usual density and redshift-space distortion terms. The third term comes from evaluating the zero-order brightness temperature at the perturbed time corresponding to the observed redshift. The time perturbation given in Eq. (14) contains integrated Sachs-Wolfe (ISW), potential and Doppler terms. The fourth term arises from the part of the ISW term in Equation (13) that is not cancelled by the velocity evolution introduced via the Euler equation. The final potential term comes from the conversion between increments in redshift and radial distance in the gas frame.

The perturbation to the brightness temperature depends only on the observer velocity through $\delta\eta$. Isolating these terms gives

$$\Delta_{T_b}^{\text{v}_{oA}}(z, \hat{\mathbf{n}}) = \hat{\mathbf{n}} \cdot \mathbf{v}_{oA} + \hat{\mathbf{n}} \cdot \mathbf{v}_{oA} (1 + z) \frac{d \ln \bar{T}_b}{dz}. \quad (18)$$

We compare this dependence with the exact result for the variation in the brightness temperature under a change of frame: $(1 + z')T'_b(z', \hat{\mathbf{n}}') = (1 + z)T_b(z, \hat{\mathbf{n}})$ which follows from the invariance of the distribution function. Since $1 + z = (1 + z')(1 + \hat{\mathbf{n}} \cdot \mathbf{v}_{\text{rel}})$ to linear order, where \mathbf{v}_{rel} is the relative velocity of the two observers, Taylor expanding the redshift dependence of $T_b(z, \hat{\mathbf{n}})$ about z' we have

$$\Delta'_{T_b}(z', \hat{\mathbf{n}}') \approx \Delta_{T_b}(z', \hat{\mathbf{n}}) + \hat{\mathbf{n}} \cdot \mathbf{v}_{\text{rel}} + \hat{\mathbf{n}} \cdot \mathbf{v}_{\text{rel}} (1 + z') \frac{d \ln \bar{T}_b}{dz'}, \quad (19)$$

which is clearly consistent with the velocity dependence in Eq. (18).

The large-scale clustering of the HI gas follows that of the discrete sources (e.g. DLAs) housing the neutral gas. We assume linear scale-independent bias of these sources in the comoving gauge, following the arguments in [27]. During matter domination, the comoving gauge coincides with the synchronous gauge, and we have

$$\delta_n = b\delta_m^{\text{syn}} + \left(\frac{d \ln(a^3 \bar{n}_{\text{HI}})}{d\eta} - 3\mathcal{H} \right) \frac{v}{k}, \quad (20)$$

where v is the Newtonian-gauge matter/gas velocity with $\mathbf{v} = -k^{-1}\nabla v$, and δ_m^{syn} is the synchronous-gauge matter overdensity.

Equation (17) can be compared with the related expression for the differential number counts of discrete objects detailed in [26–28, 33]. For simplicity, consider counting the HI atoms per solid angle and per redshift $n_{\text{HI}}^{\text{obs}}(z, \hat{\mathbf{n}})$. The brightness temperature, Eq. (6), may be rewritten as

$$T_b(z, \hat{\mathbf{n}}) = \frac{3h^3 A_{10}}{32\pi k_B E_{21}^2} \frac{n_{\text{HI}}^{\text{obs}}(z, \hat{\mathbf{n}})}{\det \mathcal{D}}, \quad (21)$$

³ We retain the local dipole term at the observer for completeness, but, since this is observer-dependent at linear order, we only consider multipoles $l \geq 2$ in our later forecasts.

⁴ Note that this equation is still valid in the modified-gravity models we are interested in.

where $\det\mathcal{D}$ is the determinant of the Jacobi map in the observer's frame [34] which relates the (Lorentz-invariant) area of a bundle of light rays to the solid angle at the observer. The determinant of the Jacobi map is related to the square of the luminosity distance d_L through reciprocity: $d_L^2 = (1+z)^4 \det\mathcal{D}$. The presence of d_L^2 in Eq. (21) makes intuitive sense since we measure surface brightness (i.e. angular density of sources multiplied by flux) in 21 cm intensity mapping experiments rather than counting objects. At linear order the fractional perturbation to the brightness temperature $\Delta_{T_b}(z, \hat{\mathbf{n}})$ is then related to the HI angular number density fluctuation $\Delta(z, \hat{\mathbf{n}})$ (calculated following the galaxy counts calculation in [26–28, 33]) by

$$\Delta_{T_b}(z, \hat{\mathbf{n}}) = \Delta(z, \hat{\mathbf{n}}) - 2 \frac{\delta d_L(z, \hat{\mathbf{n}})}{\bar{d}_L(z)}, \quad (22)$$

where \bar{d}_L is the background luminosity distance and δd_L denotes its first-order perturbation. The brightness temperature is therefore a combination of angular number density fluctuations and luminosity distance fluctuations. In particular, Eq. (22) tells us why the brightness temperature has no magnification term at linear order. The perturbed luminosity distance at fixed observed redshift indeed contains a lensing convergence term due to transverse distortion of the volume element at the source [35]. However, so does the differential number count since we observe the *density* of a quantity which thus involves the perturbed volume element. These two effects exactly cancel at linear order – magnified sources give a greater observed flux but have reduced angular density on the sky – and lensing therefore conserves surface brightness. Hence, as for the CMB, gravitational lensing has no effect at first order to the 21 cm brightness temperature: the lensed image of a smooth sky is a smooth sky⁵.

The new terms that have been uncovered by this self-consistent linear analysis are expected to be negligible on all but the largest scales and highest redshifts (see [26] for a thorough investigation in the context of galaxy surveys). The only new aspect of our application is the narrow window function in redshift-space which enhances the relative importance of the redshift-space distortion term. Considering fluctuations at a given small angular scale $1/l$, if the redshift window function is broad compared to the typical linear scale $\chi(z)/l$ of the contributing perturbations (where $\chi(z)$ is comoving distance back to redshift z), we are in the ‘Limber’ regime and the redshift distortions $\propto -\hat{\mathbf{n}} \cdot (\hat{\mathbf{n}} \cdot \nabla \mathbf{v})$ will tend to cancel out on integrating through the window. Normalising the window function to unity to keep the integrated background \bar{T}_b almost constant, the power from redshift-space distortions falls as $(\Delta\chi)^{-2}$, where $\Delta\chi$ is the width of the window function, on small scales. In the Limber limit, the power from the integrated density term falls more slowly – as $(\Delta\chi)^{-1}$ – since the fluctuation *power* accumulates as the number of incoherent patches within the window. However, since the redshift window is very narrow for 21 cm mapping, only very small scales are in the Limber regime, thus significantly enhancing the relative power in redshift-space distortions, as seen in Fig. 6 of [26].

In Fig. 1 we plot the auto-spectra (see Sec. II A for details) of each of the terms in Eq. (17) individually, including those contained in $\delta\eta$, for a bandwidth of 2 MHz at $z = 1$. Clearly, the density and redshift-space distortion terms are dominant on all angular scales at this redshift. Note that super-Hubble effects are generally suppressed in the 21 cm power spectrum since the signal on large angular scales is dominated not by modes at the corresponding linear scale, but by smaller sub-Hubble scale modes [29]. This is because the dimensionless power spectrum of δ_n grows rapidly as k^4 on scales smaller than Hubble length (but larger than the horizon size at matter-radiation equality). The 21 cm signal is therefore like white noise, $C_l = \text{const.}$, on angular scales large compared to the angle subtended by the peak in the matter power spectrum.

The fractional error in the power spectrum if only the density⁶ and redshift-space distortion terms are retained is shown in Fig. 2 at $z = 1$ for various widths of redshift window. We see that the relative importance of relativistic effects increases as the bin size increases, consistent with the results of [26]. This arises from the dominant white-noise contribution of small scale modes at a given l being suppressed by cancellations through the width of the window, but the contribution of large-scale modes, where the relativistic terms are relevant, not being suppressed. At the low redshifts considered in this work, the relativistic terms represent only sub-percent corrections to the large-scale angular power, where cosmic variance is large. Moreover, their effect is small compared to astrophysical modelling uncertainties, for example in the bias. However, we include the relativistic terms in our forecasts for consistency.

⁵ Note, however, that at second order 21 cm intensity mapping can be used to reconstruct the lensing power spectrum [24].

⁶ In the remainder of this work, ‘density’ refers to the Newtonian-gauge density of Eq. (20).

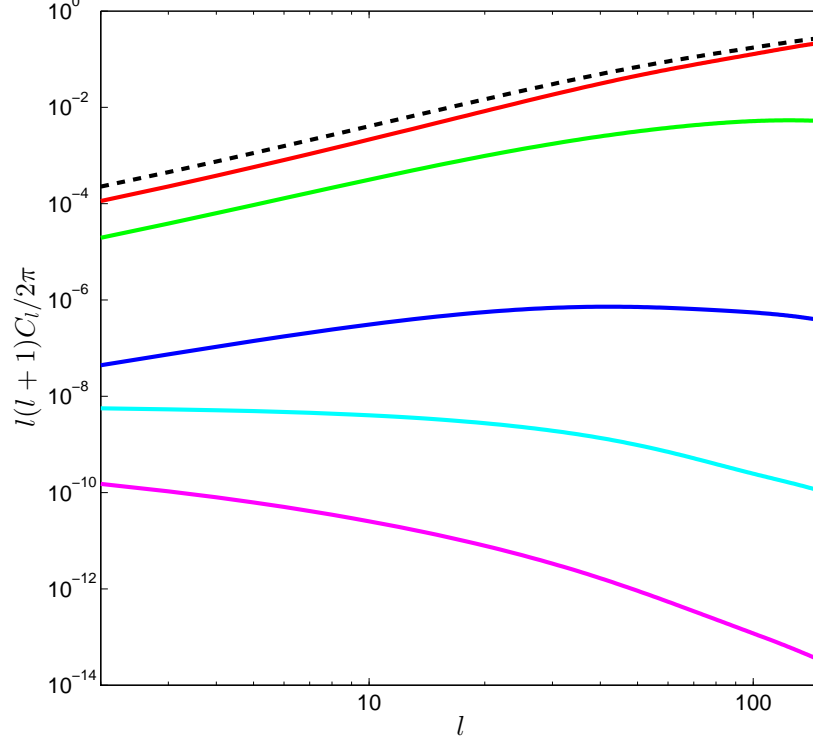


FIG. 1: (Color Online). Fractional brightness temperature perturbation power spectrum at $z = 1$ with a 2 MHz bandwidth. The auto-spectra of the full signal (black, dashed) and of each individual term in Eq. (17) are shown, generically grouped (solid lines, top to bottom respectively) as Newtonian-gauge density (red), redshift-space distortions (green), velocity terms (blue), all potential terms evaluated at the source position (cyan) and the ISW term (magenta).

A. Angular power spectra

After a Fourier transform and expansion into angular multipoles we have the brightness temperature perturbation in k and l space. For example, the first two terms in Eq. (17) give

$$\Delta_{T_b,lm}(z) = 4\pi i^l \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \Delta_{T_b,l}(\mathbf{k}, z) Y_{lm}^*(\hat{\mathbf{k}}), \quad (23)$$

where

$$\Delta_{T_b,l}(\mathbf{k}, z) \approx \delta_n j_l(k\chi) + \frac{kv}{\mathcal{H}} j_l''(k\chi), \quad (24)$$

with χ the conformal distance. We integrate over a redshift (frequency) normalized window function $W(z)$ to give

$$\Delta_{T_b,l}^W(\mathbf{k}) = \int_0^\infty dz W(z) \Delta_{T_b,l}(\mathbf{k}, z). \quad (25)$$

The angular cross-spectra between redshift windows is then calculated as

$$C_l^{WW'} = 4\pi \int d\ln k \mathcal{P}_{\mathcal{R}}(k) \Delta_{T_b,l}^W(k) \Delta_{T_b,l}^{W'}(k), \quad (26)$$

where $\mathcal{P}_{\mathcal{R}}(k)$ is the dimensionless power spectrum of the primordial curvature perturbation \mathcal{R} , and the transfer functions $\Delta_{T_b,l}^W(k) \equiv \Delta_{T_b,l}^W(\mathbf{k})/\mathcal{R}(\mathbf{k})$.

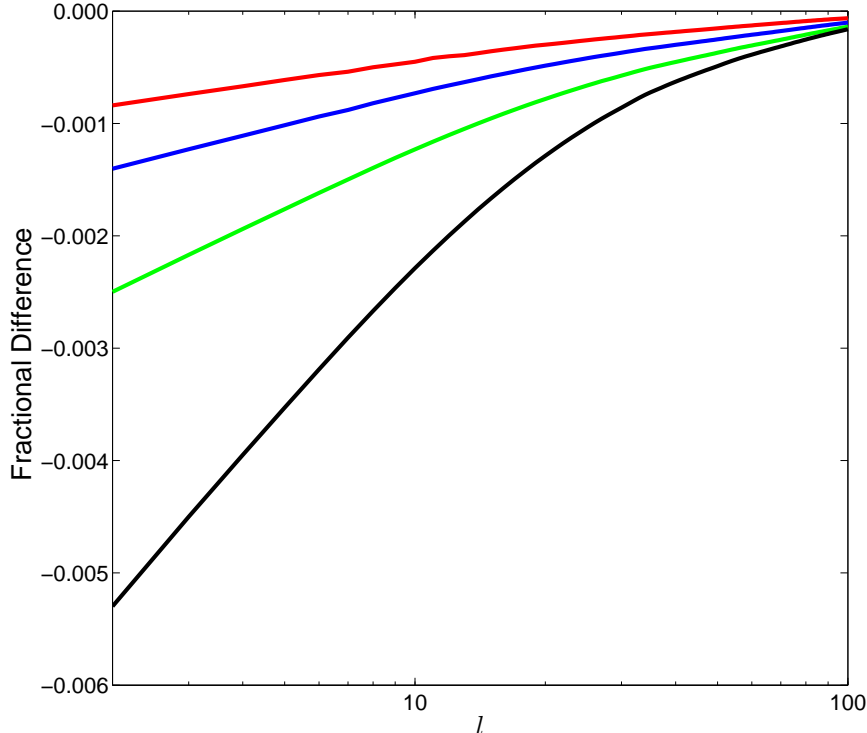


FIG. 2: (Color Online). Fractional error in the 21 cm angular power spectrum at $z = 1$ if only density and redshift-space distortions terms are retained. The errors are plotted for window functions of bandwidth (top to bottom respectively) 2 MHz (red), 5 MHz (blue), 10 MHz (green), and 20 MHz (black).

We calculate the cross-spectra using a modified version of the Boltzmann code CAMB SOURCES⁷. At the low redshifts we consider, some care must be taken when integrating over the narrow window function, and we found it necessary to run the code at `accuracy_boost = 2` to ensure the window was well sampled.

III. MODIFIED GRAVITY

Several attempts have been made to explain the phenomenon of accelerating expansion through a modification to standard GR [3]. These modifications must predict an expansion history close to that of Λ CDM, but generically predict different linear perturbation dynamics [4, 36, 37]. It is therefore necessary to study the clustering of matter in order to distinguish between competing theories⁸.

Amongst the most studied examples of modified gravity are scalar-tensor theories and the higher-derivative theory $f(R)$, which can be mapped on to a scalar-tensor theory via a conformal transformation of the metric and a field redefinition [38]. The action in a scalar-tensor theory takes the form

$$S = \int d^4x \sqrt{-g} \left[-\frac{M_{\text{Pl}}^2}{2} R + \frac{1}{2} (\nabla\phi)^2 - V(\phi) \right] + S_m(\tilde{g}_{\mu\nu}, \psi_m^{(i)}), \quad (27)$$

where $M_{\text{Pl}} = 1/\sqrt{8\pi G}$ is the reduced Planck mass, V is the potential for the scalar field ϕ , and $\psi_m^{(i)}$ are the matter

⁷ <http://camb.info/sources>.

⁸ It should be borne in mind that none of these theories explains why the vacuum energy from particle physics is cancelled to such high precision on cosmological scales.

fields which couple to the conformally rescaled metric $\tilde{g}_{\mu\nu}$ where

$$\tilde{g}_{\mu\nu} = e^{-\alpha(\phi)/M_{\text{Pl}}} g_{\mu\nu}, \quad (28)$$

where $\alpha(\phi)$ is an arbitrary coupling function. The metric $g_{\mu\nu}$ is the Einstein-frame metric, where the action in Eq. (27) looks like the standard Einstein-Hilbert action but with matter non-minimally coupled to the metric. This frame has the advantage of being mathematically ‘close’ to GR, but has the disadvantage that matter does not follow the geodesics of $g_{\mu\nu}$ and the energy-momentum tensor is not covariantly conserved with respect to this metric. The frame picked out by $\tilde{g}_{\mu\nu}$ is the Jordan frame, where matter moves along geodesics but the gravitational action is not of Einstein-Hilbert form. Both frames are equivalent in the sense that observables calculated in either will be the same. From now on we will assume that all matter fields couple to the metric with a universal coupling $\alpha(\phi)$.

We describe linear perturbations in the conformal Newtonian gauge. For scalar modes, the field equations reduce to four independent equations involving the gravitational potentials Φ and Ψ , the fractional density perturbation δ , the velocity potential v (or equivalently the momentum density q), and the anisotropic stress Π . The anisotropic stress is negligible in the late Universe, but we include it in numerical work for consistency. In the Jordan frame, energy-momentum is conserved and so the continuity and Euler equations retain their forms in GR. In Fourier space, for pressure-free matter, they become

$$\dot{\delta} + kv - 3\dot{\Phi} = 0, \quad \dot{v} + \mathcal{H}v - k\Psi = 0, \quad (29)$$

where the dots denote derivatives with respect to conformal time η . We use the Fourier convention

$$\Psi(\mathbf{x}, \eta) = \int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \Psi(\mathbf{k}, \eta) e^{i\mathbf{k}\cdot\mathbf{x}}, \quad (30)$$

and, recall, $v(\mathbf{k}, \eta)$ is defined through $\mathbf{v} = -i\hat{\mathbf{k}}v$. This leaves two independent equations that may receive modifications. It is convenient to take these to be the Poisson equation, and the relationship between the two potentials. In many models, including scalar-tensor theories, the modifications on sub-Hubble scales (i.e. the quasi-static limit where temporal derivatives are small compared to spatial derivatives) take the form

$$-k^2\Psi = 4\pi G a^2 \mu(a, k) \bar{\rho} \Delta, \quad (31)$$

$$\Phi = \gamma(a, k) \Psi, \quad (32)$$

where Δ is the comoving density perturbation $\Delta = \delta + 3\mathcal{H}v/k$ and $\bar{\rho}$ is the background matter density. Equations (31) and (32) may be taken to define the functions μ and γ which are generally functions of time and scale; in GR, $\mu = \gamma = 1$. Note that when the quasi-static limit is not obtained, we may still define μ and γ through Eqs(31) and (32) but these functions will now generally depend on the initial conditions. Note that, for later convenience, we have written the modified Poisson equation in terms of Ψ whereas it involves Φ in GR.

It is straightforward to show that, despite the modifications in Eqs (31) and (32), the comoving curvature perturbation $\mathcal{R} = -\Phi - \mathcal{H}v/k$ is conserved on super-Hubble scales as it is in GR. Specifically, we find

$$\left[3\frac{H'}{H} - \left(\frac{k}{aH} \right)^2 \right] \mathcal{R}' = \left(\frac{k}{aH} \right)^2 (\Phi' + \Psi) + \frac{k^2 H'/H}{4\pi G a^2 \mu \bar{\rho}} (\Psi' - [\ln(a^2 \mu \bar{\rho})]' \Psi), \quad (33)$$

where $H = \mathcal{H}/a$ is the Hubble parameter and primes denote derivatives with respect to $\ln a$. Since $\mathcal{R} \sim \Phi$, the comoving curvature perturbation is conserved on super-Hubble scales provided μ and γ tend to scale-free functions there. Constancy of the comoving curvature perturbation can be shown to hold very generally for metric theories of gravity in which energy-momentum is conserved [39, 40]. Differentiating $\mathcal{R}' = 0$ gives a second-order equation for the potentials which in our notation is

$$\Phi'' + \Psi' - \frac{H''}{H'} \Phi' + \left(\frac{H'}{H} - \frac{H''}{H'} \right) \Psi = 0 \quad (k \ll aH). \quad (34)$$

Note that, once the relation between Φ and Ψ is specified, the evolution of the metric perturbations on super-Hubble scales is determined by the background expansion history [40]. For matter-dominated expansion, $H'/H = H''/H'$ and Φ and Ψ are constant on large scales if γ is time-independent. We are, of course, still free to modify the relationship between Ψ and Δ ; for example, $\mu > 1$ on large scales with constant γ would actually *suppress* large-scale clustering in Δ during matter-dominated expansion, despite increasing μ giving a larger effective Newton’s constant.

On sub-horizon scales, we can rewrite the set of perturbation equations as second-order evolution equations for the density and for the velocity. Neglecting the $\dot{\Phi}$ term in the continuity equation and replacing Δ with δ in the Poisson equation, we find

$$\ddot{\delta} + \mathcal{H}\dot{\delta} - 4\pi G a^2 \mu \bar{\rho} \delta = 0, \quad (35)$$

$$\ddot{v} + \left(2\mathcal{H} - \frac{\dot{\mu}}{\mu}\right) \dot{v} + \left(\dot{\mathcal{H}} + \mathcal{H}^2 - \mathcal{H}\frac{\dot{\mu}}{\mu} - 4\pi G a^2 \mu \bar{\rho}\right) v = 0. \quad (36)$$

We see that the evolution of δ and v on sub-horizon scales is completely determined by μ . The density and redshift-space distortion terms in the 21 cm brightness temperature are therefore insensitive on small scales to modifications in the γ function. However they are expected to deliver stringent constraints on the time and scale dependence of μ . In particular, since the evolution of the velocity is sensitive to both μ and its time derivative $\dot{\mu}$, it is strongly affected by sharp variation in μ .

To understand qualitatively the effect of μ on the growth of perturbations on sub-Hubble scales, consider a toy-model with constant μ and matter-dominated expansion. Equation (35) then becomes

$$\ddot{\delta} + \frac{2}{\eta}\dot{\delta} - \frac{6\mu}{\eta^2}\delta = 0, \quad (37)$$

which has power-law solutions $\delta \propto a^{p_{\pm}}$ with $p_{\pm} = (-1 \pm \sqrt{1 + 24\mu})/4$. As expected, $\mu > 1$ enhances clustering and δ grows faster than in GR (where $\delta \propto a$) and the gravitational potential wells of Ψ deepen in time.

Although the above solutions were derived under the simplifying assumptions of a constant μ and sub-horizon scales during matter domination, the general effect of $\mu > 1$ on the clustering of matter is clear: a larger effective Newton's constant boosts the strength of gravity which enhances the degree of clustering on sub-Hubble scales, thus allowing the gravitational potentials to grow with time. The time-dependence of Ψ and Φ can source an ISW effect in the CMB temperature anisotropy, and alter the strength of gravitational lensing effects as a function of redshift.

Since the models we consider attempt to explain the late-time acceleration of the universe, they only affect the linear dynamics at late times when the acceleration sets in⁹. The CMB is thus only sensitive to modified gravity at the perturbative level via the large-scale (late-time) ISW effect, and the CMB weak lensing signal. Both of these are sensitive to the Weyl potential, which is equal to $(\Phi + \Psi)/2$ in the Newtonian gauge. The Weyl potential is related to the matter density by

$$-k^2(\Phi + \Psi) = 4\pi G a^2 \mu (1 + \gamma) \bar{\rho} \Delta = 8\pi G a^2 \Sigma(a, k) \bar{\rho} \Delta, \quad (38)$$

where we have defined the new parameter

$$\Sigma \equiv \frac{\mu}{2}(1 + \gamma). \quad (39)$$

We will work in the (μ, Σ) basis since the 21 cm signal is primarily sensitive to μ and the CMB is primarily sensitive to Σ , although the CMB is also sensitive to μ through the enhanced growth of Δ it induces.

Under fairly weak assumptions about the coupling function $\alpha(\phi)$ [41], an action of the form of Eq. (27) leads to the Bertschinger-Zukin [42] forms of μ and γ in the Jordan frame given by

$$\mu(a, k) = \frac{1 + \beta_1 \lambda_1^2 k^2 a^s}{1 + \lambda_1^2 k^2 a^s}, \quad (40)$$

$$\gamma(a, k) = \frac{1 + \beta_2 \lambda_2^2 k^2 a^s}{1 + \lambda_2^2 k^2 a^s}, \quad (41)$$

where β_1 and β_2 are coupling parameters, λ_1 and λ_2 are Compton wavelengths at the present epoch ($a = 1$) with s a constant describing their time evolution. For scalar-tensor theories, only two of β_i and λ_i are independent [41]. This form for μ and γ has the advantage of being physically transparent; on scales much larger than the Compton wavelength of the scalar field μ and γ obtain their GR forms, and on scales below the Compton wavelength the modifications are

⁹ There are exceptions, for example $f(R)$ models with $d^2 f/dR^2 < 0$ have unstable perturbations at high curvature for a standard expansion history [4].

proportional to the coupling parameters β_i . In addition, for $s > 0$, the comoving Compton wavelength grows with time, such that the influence of the scalar field perturbation can be confined to late times if desired.

The disadvantage of using the parametrizations in Eqs (40) and (41) is that their domain of applicability is limited to scalar-tensor theories, or theories conformally equivalent. A more general, model-independent approach is desirable. The works of Baker et al. [43, 44] provide just such a model-independent formalism, but the large number of free functions present in their formalism is undesirable for our simple application here.

We argue that the most interesting theory that may be probed by upcoming large-scale structure surveys is the *null* theory, i.e. we seek to test whether GR is correct on cosmological scales. To establish the veracity of this statement is a far more interesting challenge than choosing between the many alternative theories on offer, few of which have a compelling observational justification for pursuit. To this end, we will perform a principal component analysis (PCA) of the functions μ and Σ . Since GR predicts that these functions are equal to unity at all times and (linear) scales, any evidence to the contrary would be particularly interesting. A study of the well-constrained functional forms provides a model-independent way of forecasting future constraints¹⁰. We outline the details of our PCA procedure in Sec. V.

We note in closing that the parametrization of Eqs (40) and (41) is particularly ill-suited to inference when the maximum-likelihood point is close to GR. In this case, one of the parameters (β_i, λ_i) is unconstrained. To see this, consider fixing the coupling to zero. Then, it does not matter what value we pick for the Compton wavelength or its time dependence since we always recover GR. Similarly, if the Compton wavelength is unobservably small, we expect no constraint on the coupling or time dependence, since the effects of the scalar field are never felt at observable scales. The posterior is then very non-Gaussian in these parameters around the GR model and an error analysis based on the Fisher matrix will be wrong. A better approach would be to use some related parametrization which uses specific values of μ and γ at predefined scales and times, which should have a much more Gaussian posterior. One could then map the Fisher constraints onto non-Gaussian distributions for the (β_i, λ_i) parameters.

IV. INTENSITY MAPPING SURVEY ASSUMPTIONS

In this section we provide details of the window functions and survey parameters of our assumed 21 cm experiment, and the astrophysical bias. Throughout, we use a Gaussian frequency-space window function normalized to unity with a fixed bandwidth of 2 MHz. We consider the frequency range 400–800 MHz corresponding to $z = 0.7$ –2.5. The corresponding radial resolution is 12 Mpc at $z = 0.7$ and 34 Mpc at $z = 2.5$. Smaller bandwidths are possible and may improve our constraints somewhat since at the upper end of the redshift range we include linear modes short enough to be suppressed by integration through the window (see Sec. V). However, using a narrower window requires finer sampling and increases the run-time of the code significantly.

We take a scale-independent bias of $b = 2$, which is consistent with the bias of DLAs recently derived from cross-correlation with the Ly α forest [45]. We further take the bias to be independent of redshift. In our forecasts, we later treat the bias as a free parameter and marginalise over it, arguing that its exact numerical value is not important.

For our experimental setup, we consider a small-scale interferometer consisting of an array of dipole antennae. We take the dimensionful thermal noise angular power spectrum to be white with [46]

$$\frac{l(l+1)C_l^N}{2\pi} = \frac{T_{\text{sys}}^2(2\pi)^2}{\Delta\nu t_o f_{\text{cover}}^2} \frac{l(l+1)}{l_{\text{max}}^2}, \quad (42)$$

and the noise uncorrelated between frequency bins. Here, T_{sys} is the system temperature of the antennae, $\Delta\nu$ is the bandwidth, t_o is the total observing time, f_{cover} is the ratio of the total collecting area to the geometric area of the array, and l_{max} is the maximum multipole given by $2\pi D/\lambda$ where λ is the observing wavelength and D is the diameter of the array. The quantity f_{cover} is sometimes called the array filling factor in the literature, and is closely related to the aperture efficiency of a single-dish experiment. We set $f_{\text{cover}} = 1$ for simplicity, since its precise value depends on the geometry of the experiment. The system temperature is given by $T_{\text{sys}} = T_{\text{ant}} + T_{\text{sky}}$ where T_{ant} is the antenna temperature and T_{sky} is the angle-averaged sky temperature due to foreground contamination. At the low redshifts considered here, the system temperature is dominated by the antenna temperature, which we take to be 40 K, a reasonable estimate given current technology [11]. For the sky temperature we take

$$T_{\text{sky}} = 5.0 \left(\frac{\nu}{710 \text{ MHz}} \right)^{-2.6} \text{ K} \quad (43)$$

¹⁰ However, since the number of PCA nodes is finite, we are still vulnerable to theories which may mimic GR at the nodes, but the use of a fine enough pixelization should mitigate this risk considerably.

TABLE I: Experimental parameters adopted for our forecasts; see text for definitions.

Parameter	Value
ν_{\max}	835 MHz
ν_{\min}	406 MHz
$\Delta\nu$	2 MHz
D	100 m
f_{cover}	1.0
t_o	1 yr
T_{ant}	40 K
f_{sky}	0.5

appropriate for synchrotron radiation. We note that Eq. (42) has been derived under the assumption that the array uniformly covers a region of Fourier space of area πl_{\max}^2 as the Earth rotates; for refinements accounting for non-uniformity in the array see e.g. [46].

We assume a total integration time of one year, and assume the array has a diameter of $D = 100$ m, corresponding to something roughly the size of CHIME [18]. In addition, we assume the survey covers a fraction of the sky $f_{\text{sky}} = 0.5$. Our experimental parameters are summarized in Table I.

Since the sources of the 21 cm signal are discrete, the measured auto-spectra have shot noise contributions as well as the clustering signal. For a redshift window with an angular density of sources $\bar{N}(z)$, we include a dimensionless shot noise power given by $C_i^{\text{shot}} = 1/\bar{N}(z)$. We assume a comoving number density of sources of $0.03 h^3 \text{Mpc}^{-3}$ following [24].

We further assume that no modes are used with multipole greater than some z -dependent cut-off $l_{\max}^{21\text{ cm}}(z)$ to avoid issues with modelling non-linear scales (see Sec. V). In Fig. 3 we plot the dimensionless 21 cm auto-spectra and noise power spectra at $z = 2.5$ and $z = 1.0$. In each case, the greatest multipole plotted is the cut-off $l_{\max}^{21\text{ cm}}(z)$.

We further split the frequency range into N_{win} 2 MHz frequency windows. The choice of N_{win} and the spacing of the windows is a subtle issue dependent to some extent on what the experiment is trying to measure. Ideally, we should include all $(\nu_{\max} - \nu_{\min})/\Delta\nu$ windows to retain all information. However, the number of power spectra increases with the number of windows as N_{win}^2 and calculating all of these is computationally prohibitive for the forecasts in this paper. This problem would be exacerbated in a full parameter analysis where the spectra would have to be computed $O(10^5)$ times across the parameter space. Of course, the observed signal will be highly correlated for sufficiently small radial separations. Fine sampling on radial scales smaller than this correlation length will not improve parameter constraints significantly in models with power spectra that vary slowly in time if the measurements in individual windows are sample-variance limited. Note that for 21 cm cosmology, on large angular scales the radial correlation length is rather smaller than $\chi(z)/l$ since shorter modes contribute significantly there. This means that even if one is interested in a parameter that only affects large scales, it is advantageous to space windows more closely than this to beat down the cosmic variance of the smaller-scale modes. An alternative approach is to transform to a different radial basis to remove the correlations. For example, if we were simply observing projections of a time-independent scalar field, a spherical-Bessel transform in the radial direction (adopting a fiducial cosmology to convert redshift to distance) would yield approximately uncorrelated spherical multipoles. Such an approach is advocated in [47] for large-scale cosmic shear surveys. Note that galaxy redshift surveys circumvent the issue on small scales by breaking the survey volume up into sub-volumes and estimating the (anisotropic) 3D power spectrum independently in each volume.

A full exploration of these issues is beyond the scope of this paper. Instead, for most of our forecasts we simply take $N_{\text{win}} = 20$ windows equally spaced across the observable frequency range. As we discuss in Sec. V A and V B, we expect our qualitative results to be stable to increases in N_{win} and to changes in their location. However, in our PCA of general models we do expect that the constraints on the amplitudes of individual modes could be improved though not their general form or the number of well-determined modes.

V. FORECASTING METHODOLOGY AND RESULTS

In this section we present the details of our statistical analysis and forecasted constraints on the parameters of modified gravity.

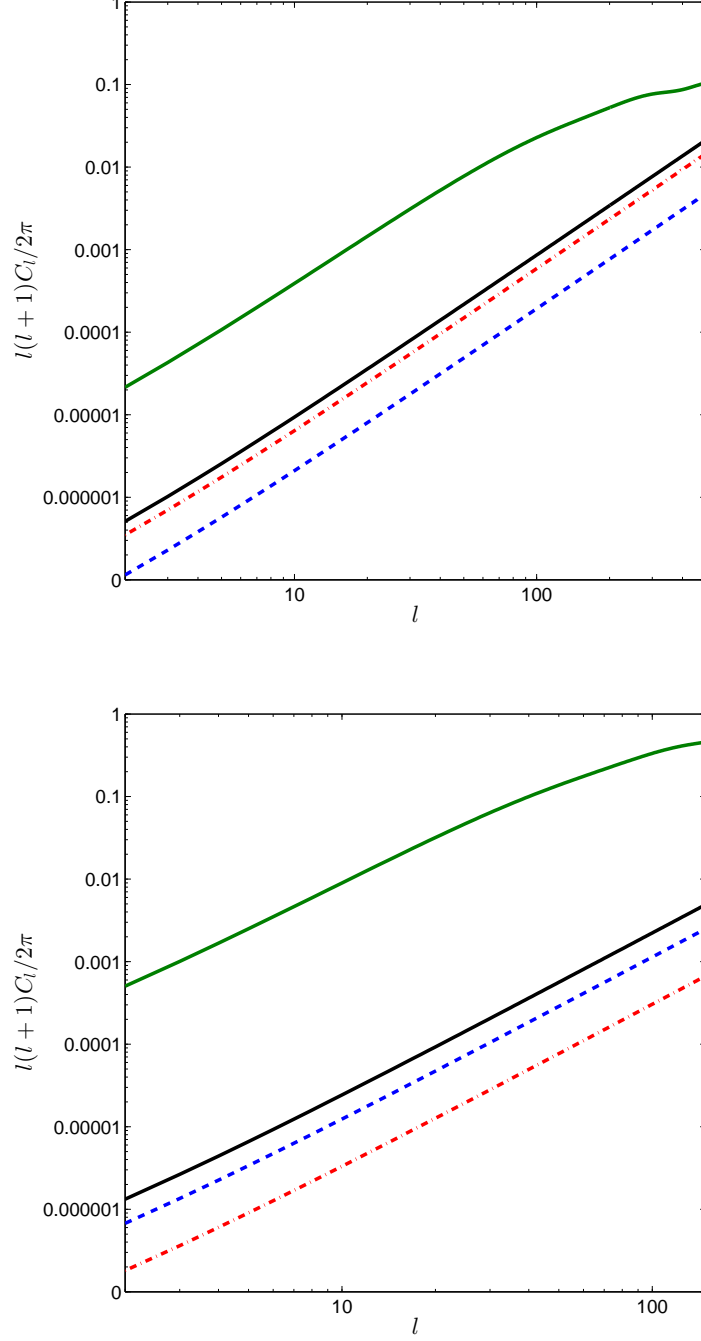


FIG. 3: (Color Online). Dimensionless 21 cm brightness temperature auto-spectra (green, upper solid), shot noise power (blue, dashed), thermal noise power (red, dot-dashed) and total noise power (black, lower solid) at $z = 2.5$ (top) and $z = 1.0$ (bottom), for the survey parameters in Table I.

We construct a Fisher matrix for the parameters θ_i (seven cosmological + N modified gravity), for an upcoming 21 cm intensity mapping experiment with noise spectrum given by Eq. (42) and survey parameters from Table I. In addition we include CMB (unlensed) temperature, E -mode polarization and a reconstruction of the lensing deflection field d from the CMB. We adopt Planck-like noise levels and angular resolutions taken from [48], and retain multipoles up to $l_{\text{max}}^{\text{CMB}} = 2000$. Statistical noise on the lensing deflection reconstruction is computed using the optimal quadratic estimator of [49].

The Fisher matrix for the parameters θ_i is defined as the ensemble average of the Hessian matrix of the log-likelihood, and, assuming Gaussian fields with covariance \mathbf{C} , is given by

$$\begin{aligned} F_{ij} &= \frac{1}{2} \text{Tr} \left[\mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_i} \mathbf{C}^{-1} \frac{\partial \mathbf{C}}{\partial \theta_j} \right] \\ &= \sum_{l=2}^{l_{\max}} \sum_{XX', YY'} \frac{\partial C_l^{XX'}}{\partial \theta_i} [\text{Cov}(XX', YY')]_l^{-1} \frac{\partial C_l^{YY'}}{\partial \theta_j}, \end{aligned} \quad (44)$$

where $\text{Cov}(XX', YY')$ is the covariance of the power spectra estimators, including noise, and XX' and YY' stand for combinations of the observed fields (TT , TE , Ed etc.). For the specific form of $\text{Cov}(XX', YY')$, see for example [50]. Note that we include all the cross-correlations between the CMB and 21 cm fields, which potentially carry a lot of information about late-time phenomena (see [51] for a discussion in the context of galaxy surveys). The non-zero cross-correlation between the low redshift 21 cm signal and the (unlensed) CMB temperature anisotropy is due to the late-time ISW effect in the CMB¹¹, whilst the 21 cm-polarization correlation arises since the last-scattering surface of an electron at reionization has some overlap with the radial distances of relevance for the 21 cm signal [52]. The latter correlation is small, but grows with redshift and we include it for consistency. The inverse of the Fisher matrix gives the covariance matrix between the parameters and its diagonal elements give the 1σ marginalised errors on parameters.

We vary seven cosmological parameters, $(\Omega_b h^2, \Omega_c h^2, h, \tau, A_s, n_s, \Omega_\nu h^2)$, and assume three degenerate neutrinos at the minimum mass of the normal hierarchy ($\Omega_\nu h^2 \approx 0.0006$). We fix the curvature to be zero, and the expansion history to be that of Λ CDM (plus massive neutrinos), as favoured by a variety of cosmological probes [53].

To avoid the issues of modelling non-linear structure formation in modified gravity, for the 21 cm signal we retain only multipoles up to $l_{\max}^{21\text{cm}}$. Whilst the comoving non-linear scale at the present epoch in GR is roughly at $k_{\text{NL}} \approx 0.1 \text{ hMpc}^{-1}$, in modified gravity it can be rather larger. For example, N -body simulations of $f(R)$ models in [54] show that linear physics is only recovered for $k < 0.06 \text{ hMpc}^{-1}$ at $z = 0$. To account for these effects, we set $l_{\max}^{21\text{cm}} = 150$ for $z < 1.5$, corresponding to $k_{\text{NL}}^{21\text{cm}} \approx 0.05\text{--}0.08 \text{ hMpc}^{-1}$, which roughly matches the non-linear scales found in [54]¹². For $z > 1.5$, we set $l_{\max}^{21\text{cm}} = 500$, corresponding to $k_{\text{NL}}^{21\text{cm}} \approx 0.12\text{--}0.15 \text{ hMpc}^{-1}$, since the comoving non-linear scale is still expected to grow with time in modified gravity models. We ignore non-linearities in the CMB lensing deflection power spectrum since, at Planck noise levels, the signal-to-noise on the lensing reconstruction is expected to be very low on non-linear scales.

To include the effects of modified gravity in CAMB SOURCES, we use the MGCAMB module [55], adapted for compatibility with CAMB SOURCES. We also update the sub-horizon radiation approximations to those of the full CAMB release. The May 2011 release of CAMB SOURCES uses the old sub-horizon approximations of [56], but this formalism implicitly assumes GR. We update this part of the code to use the newer formalism of [57], which can be easily generalised to include modified gravity. We also incorporate the effects of anisotropic stress from massive neutrinos consistently in MGCAMB, although the effects are negligible at late times.

A. Constraints on $f(R)$ gravity

We start by considering the constraints that 21 cm intensity mapping can place on $f(R)$ theories. In this section, we fix the bias to a constant value of $b = 2.0$, consistent with the constraints from [45]. Note that to satisfy local tests of gravity, the $f(R)$ Compton wavelength λ at present cosmological density is restricted to $\lambda \lesssim \text{Mpc}$ [58, 59]. This limits departures from GR in $f(R)$ theories consistent with Solar System tests to non-linear scales. However, here we are not interested in constraining $f(R)$ theories per se, but rather in understanding which kind of constraints 21 cm observations could place on an (effective) theory that behaves like $f(R)$ gravity at cosmological scales, without necessarily satisfying local tests.

We consider the B_0 -parametrization of $f(R)$ gravity [60], which provides a good approximation on quasi-static scales [61]. This is closely related to the scalar-tensor type parametrization discussed in Sec. III, since $f(R)$ is conformally

¹¹ Note that the 21 cm signal also contains an ISW effect [see Eq. (14)] that is perfectly correlated with the ISW contribution to the CMB for $z > 2$. However the amplitude of this correlation is subdominant with respect to the 21 cm density-ISW cross-correlation.

¹² Note that we fix the maximum l rather than the maximum k .

equivalent to a scalar-tensor theory. The specific forms of μ and γ we consider are

$$\mu(a, k) = \frac{1}{1 - 1.4 \times 10^{-8} |\lambda/\text{Mpc}|^2 a^3} \frac{1 + 4\lambda^2 k^2 a^4/3}{1 + \lambda^2 k^2 a^4}, \quad (45)$$

$$\gamma(a, k) = \frac{1 + 2\lambda^2 k^2 a^4/3}{1 + 4\lambda^2 k^2 a^4/3}, \quad (46)$$

where the parameter $B_0 \equiv 2H_0^2 \lambda^2$ is the square of the Compton wavelength of the effective scalar degree of freedom $f_R \equiv df/dR$ (the *scalaron*) in units of the current Hubble length. The prefactor in Eq. (45) accounts for the time-dependence of Newton's constant in the Jordan-frame background equations, but is practically unity for the models we consider. It is then easy to show that $\Sigma \approx 1$ in this theory, so the relationship of the Weyl potential to the density is the same as in GR (although the density is of course different).

Our parameter space thus consists of the seven cosmological parameters, plus B_0 . We take GR as a fiducial model, i.e. $B_0 = 0$ and adopt 20 window functions equally spaced in frequency. Our 1σ errors are B_0 are 1.9×10^{-4} using only the CMB information, improving to 3.7×10^{-5} with the addition of 21 cm.

We have marginalised over all other parameters, but there is only mild degradation in the constraint as a result of this. The main covariance of B_0 is with $\Omega_\nu h^2$ (correlation coefficient of 0.4). Increasing B_0 pushes the scalaron Compton wavelength to larger scales, such that a greater range of (small) linear scales experience enhanced growth and for longer. This boosts the 21 cm signal. The main effect of light massive neutrinos on the 21 cm power spectrum is a suppression of power on scales below the horizon size at the redshift where the neutrinos become non-relativistic. The suppression is scale-dependent for scales between the free-streaming scale at the transition and the free-streaming scale at the 21 cm redshift, with smaller scales exceeding the free-streaming length later and hence suppressing the growth of structure for longer [50]. On scales smaller than the free-streaming length at the 21 cm redshift the suppression is scale-free, and for scales larger than the free-streaming length at the non-relativistic transition there is no suppression at all. This small-scale suppression can partially cancel the enhancement from increasing B_0 , giving a positive correlation.

We see that the addition of information from 21 cm intensity mapping improves the constraint on B_0 by about a factor of five over the CMB-alone case. Although this is not competitive with local constraints or forecasted constraints for the Dark Energy Survey or Large Synoptic Survey Telescope [55], it demonstrates the potential usefulness of intensity mapping in constraining $f(R)$ theories. It is interesting to note that our constraints are comparable with those in [24], who considered only the baryon acoustic oscillation signature extracted from 21 cm maps, lens reconstruction from 21 cm itself and Planck-like priors, when only linear scales are included. To map from their parametrization to ours, note that $B_0 \approx -2(n+1)f_{R_0}/(4-3\Omega_m)$ where the power-law n appears in the Hu & Sawicki model of $f(R)$ [60] that they employ. Current cosmological constraints from [62] limit $B_0 < 1.1 \times 10^{-3}$ at 95% confidence, although this constraint is driven mainly by cluster abundance data where non-linear chameleon effects not accounted for in [62] could be important. Excluding the cluster abundance data, [62] quote an upper limit of $B_0 < 0.42$ at 95% confidence. Our 95% upper limit is 7.4×10^{-5} , which represents an improvement on this current constraint by roughly four orders of magnitude.

What limits the power of intensity mapping in constraining B_0 ? The main effect of B_0 on the 21 cm power spectrum is an enhancement of power on scales below the Compton wavelength, with the enhancement increasing with time. Thus, setting aside the influence of the CMB for now, we may conjecture that the upper limit set on B_0 from 21 cm data derives from the shortest observable scale at the lowest observable redshift. In our case, this is the non-linear scale¹³, which at $z = 0.7$ is roughly $0.08 h\text{Mpc}^{-1}$. The value of B_0 which sets the comoving Compton wavelength equal to this scale at $z = 0.7$ is around 3×10^{-4} . This is only a rough upper limit, since μ and γ are not step functions, and even scales above the Compton wavelength experience some modified growth. We can therefore anticipate that as our understanding of modified-gravity scenarios in the mildly non-linear regime of structure formation improves, we will be able to exploit 21 cm (and other cosmological observables) on smaller scales thus enhancing our ability to constrain scalar-tensor theories and $f(R)$.

With some experimentation, we find that our results are stable to changes in the number of windows around $N_{\text{win}} = 20$, equally spaced across the observable frequency range. For the B_0 constraints, most of the signal comes from the lowest observed redshifts, so adding more bins does not impact the constraints significantly. Our results are also stable to moving all 20 bins to low redshift, with a radial separation corresponding to the distance at which the cross-correlation with the $z = 0.7$ window at $l = 150$ first goes to zero.

¹³ Note that our choice for the non-linear scale is more conservative than in [55]. This may partly explain their stronger constraint on B_0 .

To understand which part of the 21 cm signal is most powerful in constraining B_0 , we repeat the analysis with the density term switched off, in both cosmic variance and the signal. This is equivalent to what could be achieved if the redshift-space distortion signal could be extracted coherently in the 21 cm maps. While not possible for 21 cm intensity mapping, we note that such extraction is possible if the clustering of differently biased populations is combined [63]. We find that the marginalised error on B_0 *decreases* by roughly 60%, suggesting that redshift-space distortions are significantly more sensitive to B_0 than the density, and that the information we can extract from redshift-space distortions is limited by the cosmic variance of the dominant density term.

On the CMB side, one question we can answer quite easily is whether it is CMB lensing or the ISW which is more powerful in constraining B_0 in our combination of observables. To investigate this, we run the above analysis again but with the late ISW term set to zero. We find that the marginalised error on B_0 increases by roughly 0.5%, indicating that the information about modified gravity brought by the late ISW effect is sub-dominant to that brought by CMB lensing. This is likely due to large cosmic variance present in the CMB at the large angular scales where the late-time ISW effect is significant.

B. Principal component analysis

Whilst specific parametrizations of the modified gravity sector such as Eqs (40) and (41) have the advantage of being physically transparent and easily related to the underlying theory, they are too restrictive in scope and fail to describe the general space of theories. To test GR on cosmological scales in a more theory-agnostic manner, a model-independent approach is required whereby one considers rather general functional forms for μ and γ . Such an approach is provided by the machinery of PCA. The basic idea is to establish what functional forms of $\mu(a, k)$ and $\gamma(a, k)$ can be independently constrained well by the data. For applications of PCA in testing modified gravity, see e.g. [55, 64].

The first step in a PCA is the pixelization of the free functions, ideally at fine enough resolution that the well-determined modes coincide with what would be obtained from a continuum analysis. In practice, for this forecasting analysis, computational resources limit us to consider 4×4 grids for both μ and γ , equally spaced in both comoving wavenumber and conformal time. Our range of k is $0.00 < k < 0.153 \, h\text{Mpc}^{-1}$, and our redshift range is $0 < z < 3$. The upper limit on z is arbitrary, but we only wish to consider models where modified gravity mimics dark energy, which motivates a choice of upper limit that matches the onset of acceleration. We enforce GR for $z > 3$. The upper limit on k corresponds to the smallest observable scale in the 21 cm signal. For $k > 0.153 \, h\text{Mpc}^{-1}$ we set our functions equal to their values at $k = 0.153 \, h\text{Mpc}^{-1}$. The choice of grid resolution imposes a prior on how rapidly we allow the modified-gravity functions to vary in scale and time.

We thus parametrize departures of μ and γ from their GR form (unity) with 16 nodal values each. We use a bicubic spline to interpolate between these nodes, setting the derivatives to zero at the edges of the grid. We also experimented with bins rather than nodes, in the manner of [55, 64], but found that the steep variation of μ and γ at the bin edges caused a loss in numerical accuracy. We transform our Fisher matrix into the (μ, Σ) basis, since conditional information from Σ comes almost completely from the CMB alone, whereas information about μ comes from both 21 cm and CMB fields.

1. Fixed bias, $b = 2$

We first report results of the PCA assuming the bias $b = 2$ is known perfectly. In this case, our modified-gravity sector consists of 32 parameters and our total parameter space has a dimension of 39. We construct the Fisher matrix for all these parameters, and marginalise over the seven cosmological parameters by removing the appropriate rows and columns from the full covariance matrix. We then proceed to diagonalise this sub-matrix, which gives us 32 eigenvectors and associated eigenvalues. Each eigenvector is normalized to unity, and gives an independently constrained mode of μ and γ for our observables. The 1σ marginalised errors in the determination of the amplitude of each of these normalised modes is given by the square-root of the corresponding eigenvalue. We plot these errors in ascending order in Fig. 4.

Since the eigenmodes of the PCA refer to departures from GR, and so have fiducial value zero, there is no unambiguous definition of a ‘well-constrained’ set of principal components. However, the observations only provide useful information on those modes for which a 1σ fluctuation in amplitude corresponds to changes in μ and Σ which are not large compared to unity. Where this is not the case, physicality priors, such as $\mu > 0$, would constrain the amplitude of the mode much more strongly than the data. Moreover, it is clear that those modes for which a 1σ fluctuation in amplitude produce large changes in μ and Σ do not give a precise test of GR. Since the eigenvectors are normalised to unity in a 32-dimensional space, the typical eigenvector will have at least one component that is $O(1)$ and so we

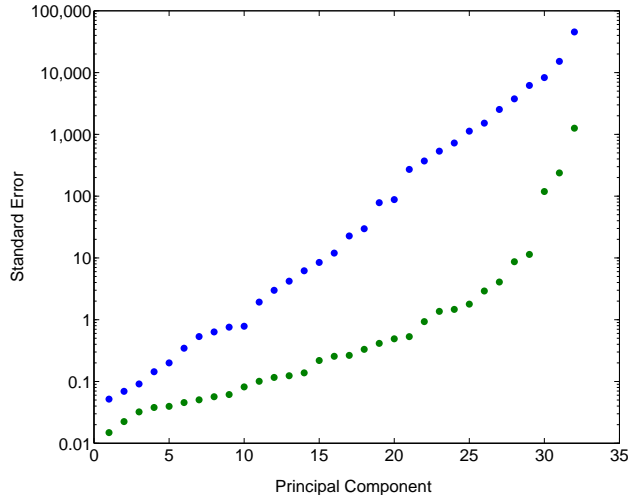


FIG. 4: (Color Online). The 1σ errors on the amplitude of each mode of the PCA (i.e. the square root of each eigenvalue) from the CMB alone (blue, upper points) and CMB and 21 cm (green, lower points).

regard eigenvalues less than one as corresponding to well-constrained modes¹⁴. On this measure, our combination of observables can constrain well around 22 modes of modifications to gravity. The discernible upturn in the spectrum of eigenvalues in Fig. 4 for the 21 cm-plus-CMB combination suggests that we have reached the saturation point of the PCA, in that the data would not be informative about the additional modes that would result from increasing the number of grid points. The upturn is at an $O(1)$ eigenvalue which makes our enumeration of the well-determined modes reasonably robust to changes in the threshold eigenvalue. We see further from Fig. 4 that adding 21 cm information to the CMB observables increases the number of well-constrained modes considerably, as well as reducing the overall errors on these modes. This is due to the information on scale and time-dependence of μ brought by the 21 cm signal, as well as degeneracy breaking effects in the combination of CMB and 21 cm, which we discuss below.

For each eigenvector, we can use its components in the (μ, Σ) basis to construct an ‘eigensurface’ in the (k, η) plane for each function, using bicubic splines to interpolate between the nodal values; see Fig. 5 for the first eight well-constrained modes and the two poorest-constrained modes. Such plots give a visual impression of the ‘sweet spots’ of μ and Σ probed by our particular observables. The poorest-constrained modes are localised to late times and small scales where our observables have little sensitivity. Note that we do not plot marginal constraints on either one of μ or Σ alone. We would expect μ and Σ to depend on functions or parameters more fundamental to the underlying theory, as is the case in scalar-tensor theories. It would thus be misleading to reduce the parameter space by marginalising over either of μ or Σ (as well as the cosmological parameters).

We see from Fig. 5 that the first two eigenmodes are dominated by Σ , both showing a single broad oscillation in the temporal direction (which is over-sampled by the grid) and confined to large scales. Although the 21 cm signal carries no information about Σ on sub-horizon scales, the information it brings on μ is complementary to that of the CMB, and breaks some of the degeneracies between μ and Σ present in the CMB. The source of this degeneracy can be understood roughly as follows: Increasing μ enhances the density perturbation on sub-horizon scales, which then increases the depth of the Weyl potential via the Poisson equation. This can be compensated by decreasing Σ , although the cancellation is not exact since the response of the density to μ is non-local in time. Since the CMB probes modified gravity only through the Weyl potential, there is an approximate degeneracy. Thus, the shapes of the Σ eigensurfaces are not the same as those that use only the CMB as an observable.

The primary sensitivity is thus to temporal oscillations in Σ , and variations in scale in μ . In Figs 6 and 7 we plot the fractional difference of the resulting power spectra from GR for the first and third eigenmodes with amplitude of 1σ . By construction, for both modes plotted the resulting change in the marginal log-likelihood is $-1/2$. Note that

¹⁴ Note that this conclusion would change if we used many more grid-points. The best determined modes are reasonably well sampled by our grid (see below) and so we expect these would not change significantly with greater grid resolution. However, the components would reduce in amplitude, to preserve the normalisation, and the eigenvalues increase to preserve the form of μ and Σ for an amplitude of 1σ . The threshold eigenvalue for well-determined modes would thus increase with grid resolution. This makes direct comparisons with other work, such as [55] which considered future cosmic shear and galaxy surveys, difficult since the dimensionality of the grids differ.

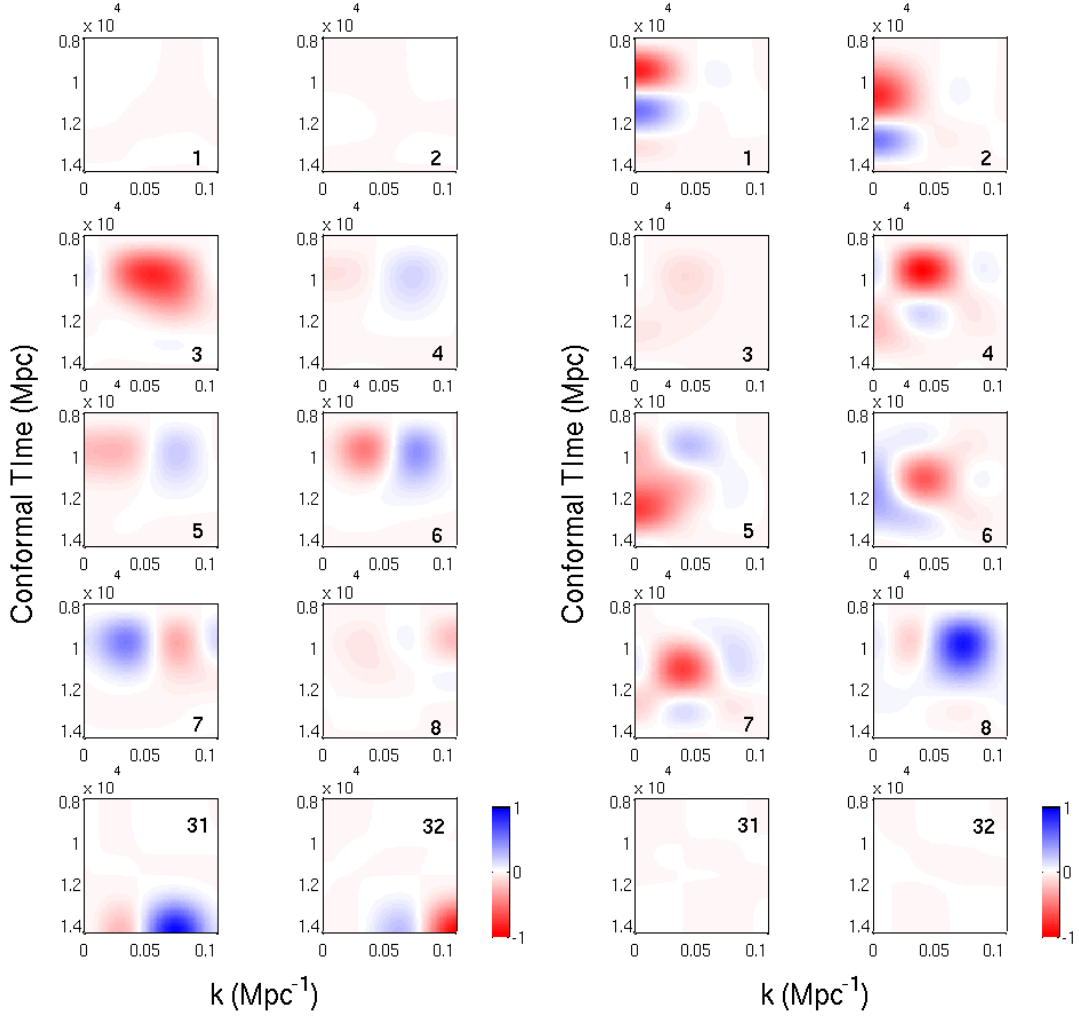


FIG. 5: (Color Online). *Left*: The $\mu(\eta, k)$ eigensurfaces of the eight most well-constrained and two poorest-constrained eigenmodes for 21 cm-plus-CMB observables in the (η, k) plane. The eigenvectors have been normalised to unity, and the surfaces are bicubic splines in between the nodes of the PCA. *Right*: The same, but for $\Sigma(\eta, k)$. Note that the overall normalisation is fixed to unity across μ and Σ but the sign of each eigenvector is arbitrary.

all differences shown are well within cosmic variance and experimental noise, but the broad-band features allow us to average over many modes and get a constraint.

We see from Figure 6 that temporal oscillations in Σ source a large ISW effect in the CMB, since the Weyl potential has a significant time derivative. There is also a lensing contribution, since the oscillations are broad enough that the CMB lensing kernel changes significantly over their period. The information on scale-dependence brought by lensing is of course limited by the Planck reconstruction noise.

To understand the shape of the lensing plot in Figure 6, consider the power spectrum of the lensing deflection d in the Limber limit, which should be accurate on small angular scales [65]:

$$C_l^{dd} \approx \frac{8\pi^2}{l} \int_0^{\chi_*} d\chi \mathcal{P}_{(\Psi+\Phi)/2}(l/\chi; \eta_0 - \chi) \frac{(\chi_* - \chi)^2}{\chi_*^2 \chi}, \quad (47)$$

where $\mathcal{P}_{(\Psi+\Phi)/2}$ is the dimensionless power spectrum of the Weyl potential, and χ_* is the conformal distance to recombination. The integral is taken along hyperbolae of constant $l = k(\eta_0 - \eta)$ in the (η, k) plane. We focus on the most well-constrained eigenmode, since this is mostly Σ with a broad oscillation in the time direction, which simplifies the discussion. In Fig. 8 we plot the Σ part of the eigensurface for this mode, with curves of constant l overlaid. Deviations from GR occur in the lensing signal because as we sum up the contributions along these curves in evaluating the integral in Eq. (47), we pass through regions where Σ does not obtain its GR value. The relative importance of these regions in the above integral depends on the magnitude of the deviation in Σ , the radial distance through the

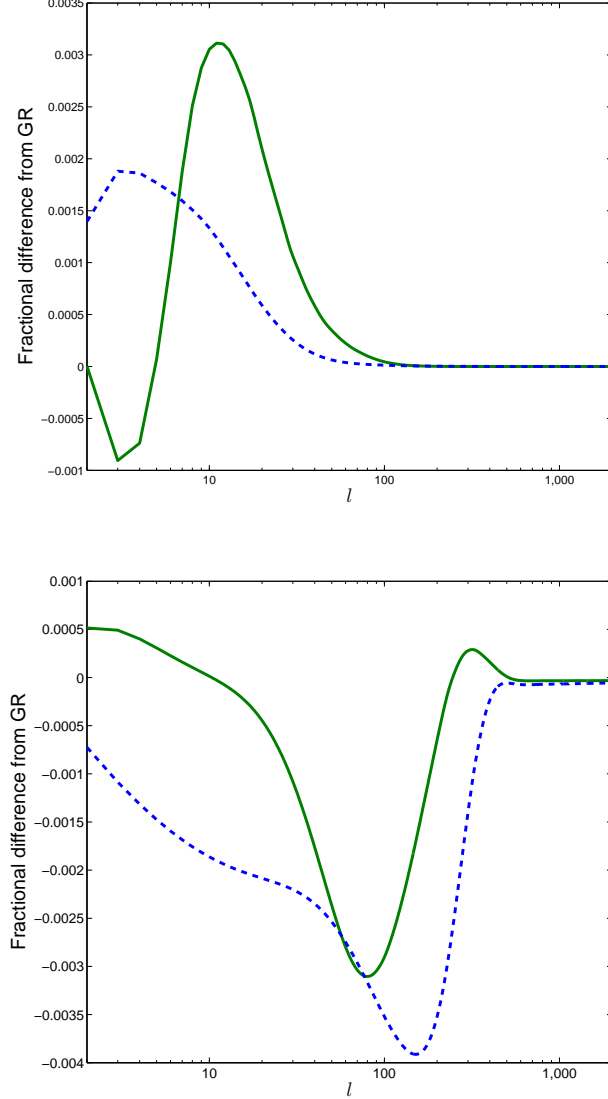


FIG. 6: (Color Online). Fractional difference from GR in the CMB temperature power spectrum (top) and CMB lensing deflection power spectrum (bottom) for 1σ changes in the first (green, solid) and third (blue, dashed) best-determined eigenmodes of the CMB-plus-21 cm combination.

region, and the product of the time-dependent geometric lensing kernel in Eq. (47) and the scale- and time-dependent value of $\mathcal{P}_{(\Psi+\Phi)/2}(k; \eta)$ in GR. The geometric kernel favours late times and the power spectrum favours large scales and early times. As an example, consider $l = 300$. The corresponding hyperbola in Fig. 8 passes through a region where $\Sigma < 1$, and a region at lower χ where $\Sigma > 1$. The geometric lensing kernel up-weights the patch at lower χ thus raising C_l^{dd} above its GR value, consistent with Fig. 6. As we decrease l , we start to pick up the large region with $\Sigma < 1$ at late times at the bottom of Fig. 8, as well as the large region in the top left corner at large scales with $\Sigma < 1$. This reduces C_l^{dd} below its GR value giving the negative feature in Fig. 6 around $l \approx 100$. At even lower l , we start to accumulate relatively more of the ‘positive’ part of the Σ oscillation, where $\Sigma > 1$. This drives the upturn in Fig. 6 at low l , although the Limber approximation become increasingly poor on these large angular scales.

In contrast, most of the well-constrained μ modes exhibit oscillations in scale, with the third most well-constrained mode having a single broad feature in the (k, η) plane. As has been mentioned, CMB lensing brings limited scale information due to reconstruction noise, and the late-time ISW effect only affects the large-scale CMB where cosmic variance is large. However, the 21 cm signal can give excellent scale information, as well as time-dependence from the tomography. Figure 7 shows the fractional difference from GR of this signal at low and high redshifts. The

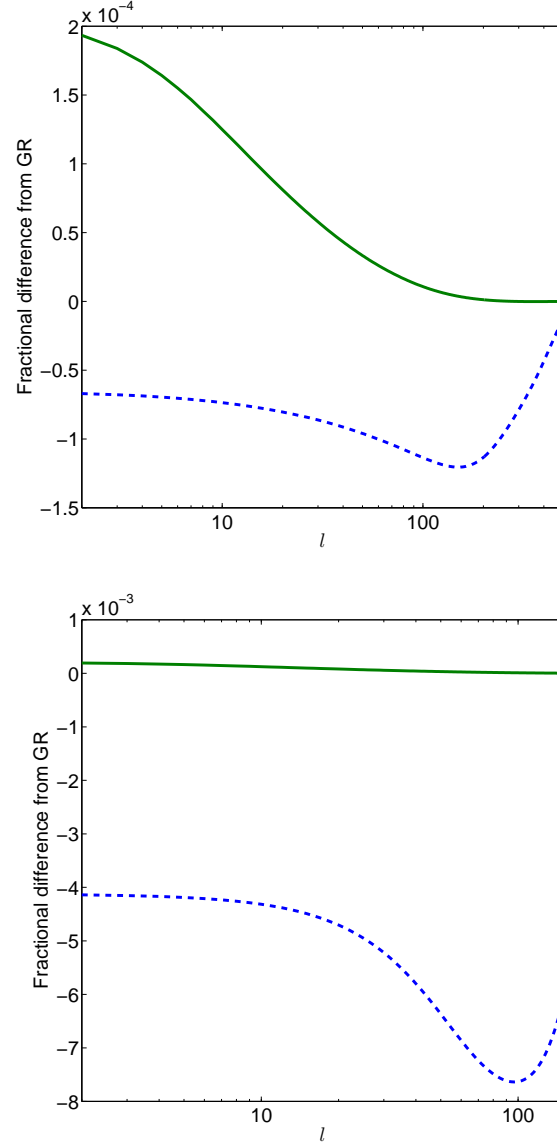


FIG. 7: (Color Online). As Fig. 6 but for the 21 cm power spectrum at $z = 2.5$ (top) and $z = 0.7$ (bottom).

well-constrained μ modes thus come mainly from the 21 cm signal, which can be confirmed by running the PCA using only 21 cm as an observable. In that case, we find the same k -space oscillations, although the overall constraints on modified gravity are poor due to degeneracies with the standard parameters which are not well constrained in the absence of CMB information.

The 21 cm power spectrum for the best-constrained eigenmode has an *excess* over GR on large angular scales at high redshift, with a magnitude larger than for the third-best constrained; see Fig. 7. This may seem surprising at first, since the $\mu(\eta, k)$ eigensurface for the best-constrained mode is low amplitude compared to the third mode and negative. However, the effects of Σ are non-negligible on horizon-scale modes in the 21 cm signal, and these purely relativistic effects drive the large deviation from GR at low l for this mode. At $l = 2$, roughly 75% of this deviation can be attributed to the new terms in Eq. (17) and their cross-correlation with the standard terms, the dominant contribution coming from the potential terms. The remaining 25% comes from non-negligible relativistic terms in the dynamical equations for the density and velocity.

To understand which term in the 21 cm signal contributes most to the μ constraints, we repeat the PCA without the density contribution. The number of principal components with eigenvalues below unity then decreases to 10. However, if we perform the PCA with bins (as in [55]), this number rises to 26, above that obtained with nodes.

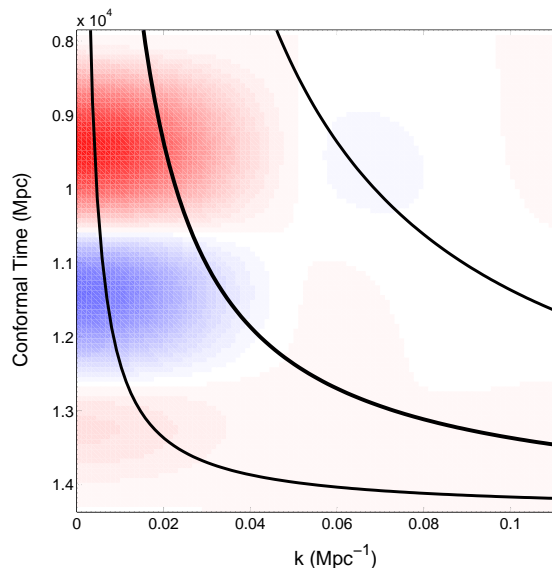


FIG. 8: (Color Online). The $\Sigma(\eta, k)$ eigensurface of the most well-constrained eigenmode for 21 cm-plus-CMB observables. The solid black curves correspond to fixed $l = k(\eta_0 - \eta)$ at values of 300 (top), 100 (middle), and 20 (bottom).

This reflects the fact that the redshift-space distortion term is significantly more sensitive to μ , and particularly to its temporal variation $\dot{\mu}$, than the density term. Hence if there were a way to separate density from redshift-space distortion in the 21 cm signal, this could significantly improve sensitivity to modifications of gravity, although uncertainties in the bias would propagate through to the variance of any density-free estimator.

To summarise, in models where the bias is constant and fixed, we are able to extract roughly 22–29 well-constrained modes from the modified-gravity sector. Our observables are most sensitive to broad radial oscillations in Σ , and broad wavenumber oscillations in μ . Addition of 21 cm information dramatically improves the number of well-constrained modes, increasing this number from 10 to 22 using the $\sigma < 1$ threshold reference. The results in this section assume $N_{\text{win}} = 20$ windows equally spaced in frequency. We speculate that increasing the number of bins could improve constraints on the individual mode amplitudes (i.e. reduce their eigenvalues) by as much as a factor of $\sqrt{10}$. However, we would not expect significant changes in the form of the eigenmodes since the 20 redshift windows already over-sample the fastest temporal variation supported by our grid. Given the sharp upturn in the spectrum of eigenvalues in Fig. 4, the number of well-constrained modes should also not change significantly with an increase in N_{win} . As a concrete (and computationally tractable) example, running the PCA with 25 equally-spaced windows does not have a significant effect on the eigenvalues of the well-constrained modes, but does improve the constraints on some of the poorly-constrained modes (although the eigenvalues of these modes still remain well above unity).

Finally, the PCA provides us with a useful consistency check. We can map μ and γ in the B_0 parametrization of $f(R)$ gravity, Eq. (46), onto the 32 nodal values and hence constrain B_0 by transforming the Fisher matrix. Doing this gives a 1σ marginalised error on B_0 of 3.5×10^{-5} , which is within 5% of the value found by using B_0 as a parameter directly.

2. Free bias parameter

Treating the bias as constant and perfectly known is unrealistic. Instead, we should model it with some functional form and marginalise over it within some prior range informed by theory and other observations. In this subsection, we treat the bias as constant – a reasonable assumption on large scales – but unknown, and marginalise over it in the forecast about a fiducial value $b = 2$. We thus expand the dimension of our total PCA parameter space to 40.

We find that the combination of 21 cm and CMB can constrain the bias with a marginalised 1σ error of 0.05, i.e. to a precision of 2.5%. As a reference, the cross-correlation of DLAs with the Lyman- α forest from [45] measures $b = (2.17 \pm 0.20)\beta_F^{0.22}$, where β_F is the Lyman- α forest redshift-space distortion parameter which is expected to be above unity. Taking $\beta_F = 1$, this represents a 9% measurement of b , but this ignores uncertainty in β_F from uncertainty in the cosmology and from modifications to gravity.

The inclusion of a constant bias as a free parameter has negligible effect on the modified-gravity constraints. The

TABLE II: Fractional increase in 1σ marginalised errors due to marginalising over a constant bias factor.

Parameter	$\sigma(\text{free } b)/\sigma(\text{fixed } b)$
$\Omega_b h^2$	1.080
$\Omega_c h^2$	1.584
h	1.148
τ	1.000
A_s	1.029
n_s	1.137
$\Omega_\nu h^2$	2.218

number of well-constrained eigenmodes in the PCA does not change. The main effect of marginalising over b is a degradation in the error on $\Omega_\nu h^2$; see Table II, where we list the ratios of marginalised errors on cosmological parameters with and without marginalising over bias. The scale dependent suppression of power induced by massive neutrinos discussed in Sec. V A can be (very approximately) mimicked by changing the bias. Even though we have forced b to be scale-independent, increasing it is not equivalent to a scale-independent amplitude change, since the redshift-space distortion part of the brightness temperature is unaffected by bias and is suppressed relative to the density term on small scales due to the width of the window function.

VI. CONCLUSIONS

We have shown that a typical near-term low-budget 21 cm intensity mapping experiment will be able to improve constraints on modified gravity theories considerably, when combined with CMB observations from Planck. In particular, this observable will improve constraints on the B_0 parameter of $f(R)$ gravity by a factor of 15 over current cosmological constraints [62].

By performing a principal component analysis on the free functions of the modified-gravity sector μ and Σ (or alternatively γ) in models with a Λ CDM expansion history, we estimate that the 21 cm experiment, combined with Planck, will provide useful information on around 20 modes. This improves on the 10 modes forecast for Planck alone, showing that 21 cm intensity mapping has significant additional constraining power in general modifications of gravity. Although our results do not seem to be competitive with those forecast in [55] for future galaxy clustering and cosmic shear surveys, our analysis is more conservative since we adopt a more realistic non-linear cut-off scale, motivated by recent N -body simulations in $f(R)$ models [54]. The 21 cm experiment is also much smaller scale and cheaper than galaxy imaging and spectroscopic surveys. More generally, the astrophysical and instrumental systematic effects are quite different and it is the combination of all such cosmological probes of modified gravity, including their cross-correlations, that will allow for accurate and precise constraints.

By studying the shape of the well-constrained eigenvectors of μ and Σ in the (η, k) plane, we can identify the ‘sweet-spots’ of the CMB-plus-21 cm data combination. We find that the combination of observables is most sensitive to broad temporal variations in Σ , and broad scale variations in μ . The former is probed primarily by the CMB, through the late-time ISW effect and gravitational lensing, although 21 cm has an important role here in breaking degeneracies with μ that exist in a CMB-only analysis. In the models we consider, with only late-time modifications to GR, the ISW effect only probes large-scale modifications. Similarly, the scale information available from CMB lensing with Planck will be limited by the high statistical noise in the lensing reconstruction, although this will improve considerably with ongoing high sensitivity polarization measurements. In contrast, the 21 cm survey provides excellent scale information which allows scale-dependent modifications in μ to be well constrained.

We have also considered the constraints that may be placed on a one-parameter model of modified gravity, specifically the B_0 parametrization often used to study $f(R)$ models. We find that our combination of observables should permit a 95% upper limit of 7×10^{-5} on B_0 after marginalising over cosmological parameters in flat models with Λ CDM expansion histories. Although this is not competitive with local constraints or forecasts from future weak lensing and galaxy surveys, it does suggest that useful constraints from cosmology can be obtained with the kind of low-budget experiments considered in this work.

Although we have considered a specific experimental set-up in this work, we have tried to keep the survey details as general as possible. In principle all of our survey parameters could be different, but we would not expect our conclusions to change significantly provided that the measurements are signal-dominated at all redshifts on scales

where linear perturbation theory is reliable (as they are in our analysis) and a significant fraction of the extra-Galactic sky is surveyed. Generally, the ratio of signal to thermal noise is best at low redshifts, and this is also where the foregrounds are lowest. Since any redshift bin can only experience the knock-on effects of modified growth that occur prior to that epoch, the low redshift bins carry a greater amount of information on modified gravity. It is precisely in these bins where the noise from the experiment is smallest.

One limitation of this work is our restriction to 20 narrow frequency windows equally spaced across the 400–800 MHz range for computational reasons. This is certainly lossy, equivalent to throwing away a significant fraction of the survey volume. While our B_0 constraint is rather stable to increasing the number of windows or using closely spaced windows at low redshift, it is not clear how sensitive our PCA results are to the same changes. Since our most well-constrained modes exhibit broad features in redshift, we do not expect the form of these modes would change, although, generally, we expect the errors on some of the mode amplitudes would improve several-fold. The total number of well-constrained modes is expected to be stable. Ways to improve our current implementation include using all $(\nu_{\max} - \nu_{\min})/\Delta\nu$ windows but only computing correlations that are expected to be significant, or, better, finding an improved radial basis that approximately decorrelates the signals. However, a detailed study of these issues is beyond the scope of this work.

A further extension to this work would be to increase the grid resolution of the PCA. Since adding more frequency windows can improve constraints on modes exhibiting grid-scale variations, we may be able to recover more well-constrained modes this way. However, it is unclear how physical rapidly varying functions in the modified-gravity sector are. We have shown that our adopted grid resolution is sufficiently fine to reproduce $f(R)$ gravity, and we speculate that the same would be true for other models. A complete Bayesian analysis would treat grids with different resolution as independent models, with the data deciding which model has the highest posterior probability, but exploring this issue is also beyond the scope of this work.

Finally, we have provided a consistent derivation of the perturbed 21 cm brightness temperature in linear theory in any metric theory of gravity. The new expression we derive includes all line-of-sight and relativistic effects, but these extra terms make only a small correction on large scales. The size of the correction increases with redshift and width of the frequency windows. We include the new terms in our analysis for numerical consistency, although they are negligible for our application to modified gravity.

Our results show that 21 cm intensity mapping provides an exciting opportunity for testing GR on cosmological scales by probing the time and scale dependence of the clustering of neutral hydrogen. Additionally, such surveys allow geometric tests of the expansion history through the baryon acoustic oscillation feature and weak lensing tomography, at a fraction of the price of other large-scale structure surveys.

VII. ACKNOWLEDGMENTS

It is a pleasure to thank Sebastien Clesse, Pedro Ferreira and Levon Pogosian for useful discussions. We especially thank Ue-Li Pen for early collaboration on this work. AH wishes to thank Philippe Brax, Elise Jennings and Richard Shaw for helpful correspondence. AH is supported by an Isaac Newton Studentship from the University of Cambridge, and the Isle of Man Government. CB is supported by the Herchel Smith Fund and by King's College Cambridge.

-
- [1] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, et al., *AJ* **116**, 1009 (1998), arXiv:astro-ph/9805201.
 - [2] S. Perlmutter, G. Aldering, G. Goldhaber, R. A. Knop, P. Nugent, P. G. Castro, S. Deustua, S. Fabbro, A. Goobar, D. E. Groom, et al., *Astrophys. J.* **517**, 565 (1999), arXiv:astro-ph/9812133.
 - [3] T. Clifton, P. G. Ferreira, A. Padilla, and C. Skordis, *Phys. Rep.* **513**, 1 (2012), 1106.2476.
 - [4] Y.-S. Song, W. Hu, and I. Sawicki, *Phys. Rev. D* **75**, 044004 (2007), arXiv:astro-ph/0610532.
 - [5] M. Kunz and D. Sapone, *Phys. Rev. Lett.* **98**, 121301 (2007), arXiv:astro-ph/0612452.
 - [6] M. Kunz, *Phys. Rev. D* **80**, 123001 (2009), arXiv:astro-ph/0702615.
 - [7] J. S. B. Wyithe and A. Loeb, *Mon. Not. R. Astron. Soc.* **383**, 606 (2008), 0708.3392.
 - [8] J. S. B. Wyithe and A. Loeb, *Mon. Not. R. Astron. Soc.* **397**, 1926 (2009), 0808.2323.
 - [9] M. McQuinn, O. Zahn, M. Zaldarriaga, L. Hernquist, and S. R. Furlanetto, *Astrophys. J.* **653**, 815 (2006), arXiv:astro-ph/0507524.
 - [10] J. R. Pritchard and A. Loeb, *Rep. Prog. Phys.* **75**, 086901 (2012), 1109.6012.
 - [11] M. F. Morales and J. S. B. Wyithe, *Annu. Rev. Astron. Astrophys.* **48**, 127 (2010), 0910.3010.
 - [12] S. R. Furlanetto, S. P. Oh, and F. H. Briggs, *Phys. Rep.* **433**, 181 (2006), arXiv:astro-ph/0608032.
 - [13] N. Dalal, O. Doré, D. Huterer, and A. Shirokov, *Phys. Rev. D* **77**, 123514 (2008), 0710.4560.
 - [14] U.-L. Pen, L. Staveley-Smith, J. B. Peterson, and T.-C. Chang, *Mon. Not. R. Astron. Soc.* **394**, L6 (2009), 0802.3239.

- [15] K. W. Masui, E. R. Switzer, N. Banavar, K. Bandura, C. Blake, L.-M. Calin, T.-C. Chang, X. Chen, Y.-C. Li, Y.-W. Liao, et al., ArXiv e-prints (2012), 1208.0331.
- [16] J. B. Peterson, R. Aleksan, R. Ansari, K. Bandura, D. Bond, J. Bunton, K. Carlson, T.-C. Chang, F. DeJongh, M. Dobbs, et al., in *Astro2010: The Astronomy and Astrophysics Decadal Survey* (2009), p. 234, 0902.3091.
- [17] R. A. Battye, M. L. Brown, I. W. A. Browne, R. J. Davis, P. Dewdney, C. Dickinson, G. Heron, B. Maffei, A. Pourtsidou, and P. N. Wilkinson, ArXiv e-prints (2012), 1209.1041.
- [18] M. Halpern, *CHIME Canadian Hydrogen Intensity Mapping Experiment*, URL <http://www.phas.ubc.ca/~halpern/chime/>.
- [19] H.-J. Seo, S. Dodelson, J. Marriner, D. McGinnis, A. Stebbins, C. Stoughton, and A. Vallinotto, *Astrophys. J.* **721**, 164 (2010), 0910.5007.
- [20] E. Visbal, A. Loeb, and S. Wyithe, *J. Cosmol. Astropart. Phys.* **10**, 030 (2009), 0812.0419.
- [21] K. W. Masui, P. McDonald, and U.-L. Pen, *Phys. Rev. D* **81**, 103527 (2010), 1001.4811.
- [22] X.-C. Mao, *Astrophys. J.* **752**, 80 (2012), 1204.1812.
- [23] T.-C. Chang, U.-L. Pen, J. B. Peterson, and P. McDonald, *Phys. Rev. Lett.* **100**, 091303 (2008), 0709.3672.
- [24] K. W. Masui, F. Schmidt, U.-L. Pen, and P. McDonald, *Phys. Rev. D* **81**, 062001 (2010), 0911.3552.
- [25] P. Brax, S. Clesse, and A.-C. Davis, ArXiv e-prints (2012), 1207.1273.
- [26] C. Bonvin and R. Durrer, *Phys. Rev. D* **84**, 063505 (2011), 1105.5280.
- [27] A. Challinor and A. Lewis, *Phys. Rev. D* **84**, 043516 (2011), 1105.5292.
- [28] J. Yoo, A. L. Fitzpatrick, and M. Zaldarriaga, *Phys. Rev. D* **80**, 083514 (2009), 0907.0707.
- [29] A. Lewis and A. Challinor, *Phys. Rev. D* **76**, 083005 (2007), arXiv:astro-ph/0702600.
- [30] J. P. Wild, *Astrophys. J.* **115**, 206 (1952).
- [31] M. A. Zwaan, M. J. Meyer, L. Staveley-Smith, and R. L. Webster, *Mon. Not. R. Astron. Soc.* **359**, L30 (2005), arXiv:astro-ph/0502257.
- [32] J. X. Prochaska and A. M. Wolfe, *Astrophys. J.* **696**, 1543 (2009), 0811.2003.
- [33] D. Jeong, F. Schmidt, and C. M. Hirata, *Phys. Rev. D* **85**, 023504 (2012), 1107.5427.
- [34] P. Schneider, J. Ehlers, and E. E. Falco, *Gravitational Lenses* (1992).
- [35] C. Bonvin, R. Durrer, and M. A. Gasparini, *Phys. Rev. D* **73**, 023523 (2006), arXiv:astro-ph/0511183.
- [36] K. Koyama and R. Maartens, *J. Cosmol. Astropart. Phys.* **1**, 016 (2006), arXiv:astro-ph/0511634.
- [37] A. Silvestri and M. Trodden, *Rep. Prog. Phys.* **72**, 096901 (2009), 0904.0024.
- [38] T. P. Sotiriou and V. Faraoni, *Rev. Mod. Phys.* **82**, 451 (2010), 0805.1726.
- [39] D. Wands, K. A. Malik, D. H. Lyth, and A. R. Liddle, *Phys. Rev. D* **62**, 043527 (2000), arXiv:astro-ph/0003278.
- [40] E. Bertschinger, *Astrophys. J.* **648**, 797 (2006), arXiv:astro-ph/0604485.
- [41] G.-B. Zhao, L. Pogosian, A. Silvestri, and J. Zylberberg, *Phys. Rev. D* **79**, 083513 (2009), 0809.3791.
- [42] E. Bertschinger and P. Zukin, *Phys. Rev. D* **78**, 024015 (2008), 0801.2431.
- [43] T. Baker, P. G. Ferreira, C. Skordis, and J. Zuntz, *Phys. Rev. D* **84**, 124018 (2011), 1107.0491.
- [44] T. Baker, P. G. Ferreira, and C. Skordis, ArXiv e-prints (2012), 1209.2117.
- [45] A. Font-Ribera, J. Miralda-Escudé, E. Arnau, B. Carithers, K.-G. Lee, P. Noterdaeme, I. Pâris, P. Petitjean, J. Rich, E. Rollinde, et al., ArXiv e-prints (2012), 1209.4596.
- [46] M. Zaldarriaga, S. R. Furlanetto, and L. Hernquist, *Astrophys. J.* **608**, 622 (2004), arXiv:astro-ph/0311514.
- [47] A. Heavens, *Mon. Not. R. Astron. Soc.* **343**, 1327 (2003), arXiv:astro-ph/0304151.
- [48] The Planck Collaboration, ArXiv Astrophysics e-prints (2006), arXiv:astro-ph/0604069.
- [49] T. Okamoto and W. Hu, *Phys. Rev. D* **67**, 083002 (2003), arXiv:astro-ph/0301031.
- [50] A. C. Hall and A. Challinor, *Mon. Not. R. Astron. Soc.* **425**, 1170 (2012), 1205.6172.
- [51] F. Schmidt, M. Liguori, and S. Dodelson, *Phys. Rev. D* **76**, 083518 (2007), 0706.1775.
- [52] A. Lewis, A. Challinor, and D. Hanson, *J. Cosmol. Astropart. Phys.* **3**, 018 (2011), 1101.2234.
- [53] E. Komatsu, K. M. Smith, J. Dunkley, C. L. Bennett, B. Gold, G. Hinshaw, N. Jarosik, D. Larson, M. R. Nolte, L. Page, et al., *Astrophys. J. Suppl. Ser.* **192**, 18 (2011), 1001.4538.
- [54] E. Jennings, C. M. Baugh, B. Li, G.-B. Zhao, and K. Koyama, *Mon. Not. R. Astron. Soc.* **425**, 2128 (2012), 1205.2698.
- [55] A. Hojjati, G.-B. Zhao, L. Pogosian, A. Silvestri, R. Crittenden, and K. Koyama, *Phys. Rev. D* **85**, 043508 (2012), 1111.3960.
- [56] M. Doran, *J. Cosmol. Astropart. Phys.* **6**, 011 (2005), arXiv:astro-ph/0503277.
- [57] D. Blas, J. Lesgourgues, and T. Tram, *J. Cosmol. Astropart. Phys.* **7**, 034 (2011), 1104.2933.
- [58] J. Wang, L. Hui, and J. Khoury, ArXiv e-prints (2012), 1208.4612.
- [59] P. Brax, A.-C. Davis, and B. Li, *Phys. Lett. B* **715**, 38 (2012), 1111.6613.
- [60] W. Hu and I. Sawicki, *Phys. Rev. D* **76**, 064004 (2007), 0705.1158.
- [61] T. Giannantonio, M. Martinelli, A. Silvestri, and A. Melchiorri, *J. Cosmol. Astropart. Phys.* **4**, 030 (2010), 0909.2045.
- [62] L. Lombriser, A. Slosar, U. Seljak, and W. Hu, *Phys. Rev. D* **85**, 124038 (2012), 1003.3009.
- [63] P. McDonald and U. Seljak, *J. Cosmol. Astropart. Phys.* **10**, 007 (2009), 0810.0323.
- [64] G.-B. Zhao, L. Pogosian, A. Silvestri, and J. Zylberberg, *Phys. Rev. Lett.* **103**, 241301 (2009), 0905.1326.
- [65] A. Lewis and A. Challinor, *Phys. Rep.* **429**, 1 (2006), arXiv:astro-ph/0601594.